

Proyecto de creación del INEE

Felipe Martínez Rizo



Textos de



Evaluación



Proyecto de Creación del INEE

Coordinación Editorial:

Miguel Á. Aguilar R.

Diana L. Flores Vázquez

Diseño y Formación:

Juan Cristóbal Ramírez Peraza

Irma Tapia Covarrubias

Instituto Nacional para la Evaluación de la Educación

José Ma. Velasco 101, Col. San José Insurgentes,

Delegación Benito Juárez, C. P. 03900, México D. F.

Primera edición, 2008

El contenido, la presentación y disposición en conjunto y de cada página de esta obra son propiedad del editor. Se autoriza su reproducción parcial o total por cualquier sistema mecánico, electrónico y otros, citando la fuente.

Impreso y hecho en México

ISBN: en trámite



Proyecto de creación del Instituto Nacional para la Evaluación de la Educación

Mayo, 2002



CONTENIDO

Presentación	7
1. Antecedentes	11
2. Justificación	17
3. Planteamiento conceptual	19
3.1. La calidad educativa y sus dimensiones	19
3.2. Las variables que influyen en la calidad educativa	20
3.3. El universo a cubrir	21
3.4. Los componentes del INEE	22
El subsistema de indicadores	24
El subsistema de pruebas	25
El subsistema de evaluación de escuelas	26
3.5. El INEE en el Sistema Nacional de Evaluación Educativa	27
3.6. La evaluación como parte de una estrategia global de calidad	31
4. Planteamiento jurídico y organizativo	35
4.1. Marco jurídico: aspectos técnicos y toma de decisiones en evaluación	35
4.2. Rasgos básicos del INEE	38
4.3. Orientaciones generales	39
4.4. Personalidad jurídica, estructura operativa y de gobierno	39
Personalidad jurídica	39
Estructura operativa	40
Estructura de gobierno	41
4.5. Financiamiento	41
5. Perspectivas de desarrollo	45
6. Conclusión	47
Referencias Bibliográficas	49
Anexo 1. Revisión de literatura sobre evaluación de sistemas educativos	51
Anexo 2. Bibliografía especializada sobre evaluación educativa	127



Presentación

Al comenzar el siglo XXI, un rasgo en que coinciden las políticas de gran número de países, de todos los niveles de desarrollo, es el relativo al elevado valor que se da a la educación, como un importante medio de mejoramiento individual y colectivo.

Por otra parte, y más allá de razonables discusiones sobre las formas apropiadas de implementarlo, también se acepta generalmente que un buen sistema de evaluación es requisito indispensable de una educación de calidad, en tanto constituye una plataforma adecuada para tomar decisiones acerca de las estrategias más convenientes para alcanzar plenamente los objetivos y fortalecer la calidad de programas y escuelas.

En consecuencia, una prioridad de la política educativa mexicana debe ser el desarrollo de un sistema nacional de evaluación, tarea de alta complejidad técnica y delicada en términos políticos, por la variedad de puntos de vista de los actores involucrados y por la importancia de sus posibles consecuencias.

Tras las elecciones del 2 de julio de 2000, el trabajo del grupo que atendió el área educativa en el Equipo de Transición del Presidente Electo, prestó atención al tema de la evaluación, por lo que el documento resultante comprendió un primer planteamiento para la creación de un organismo especializado, denominado *Instituto Nacional para la Evaluación de la Educación* (Coordinación del Área Educativa, 2000: 153-163).

El documento del Equipo de Transición, terminado en noviembre de 2000, contemplaba la creación del Instituto en un corto plazo; se solicitaron recursos para ello en el Presupuesto de Egresos para el año 2001, y el Congreso autorizó una partida para tal propósito. La preparación del proyecto completo de un organismo tan importante, sin embargo, no podía hacerse en un plazo como el que contemplaba el

documento anterior; además, en los primeros meses de la nueva administración federal, la atención debió concentrarse en la preparación del *Programa Nacional de Educación 2001-2006*, presentado al público el 28 de septiembre de 2001.

Con plena conciencia de su importancia para la calidad del Sistema Educativo Nacional, el *Programa Nacional* propone la consolidación del sistema de evaluación, como una meta prioritaria de la política educativa. Según el *Programa*, una pieza clave de tal sistema es el *Instituto Nacional para la Evaluación de la Educación* (INEE).

Desde la primera parte, en el apartado en que se resume la visión al año 2025, el *Programa* establece que en ese horizonte, los resultados del Sistema Educativo Nacional:

...serán reconocidos nacional e internacionalmente por su buena calidad, fruto del profesionalismo de los educadores, de recursos proporcionados a sus responsabilidades, del uso de la información para alimentar la planeación y la toma de decisiones, y de mecanismos rigurosos y confiables de evaluación (SEP, 2001: 71).

Además de referirse al Sistema Nacional de Indicadores, como mecanismo para saber si se avanza en dirección del cumplimiento de los propósitos del *Programa*, en la segunda parte de éste, relativa a la reforma de la gestión del Sistema Educativo Nacional, se precisa que:

...la calidad del Sistema Educativo Nacional y sus componentes se apreciará objetivamente, gracias a mecanismos rigurosos y confiables de evaluación que serán independientes de las autoridades, cuyos resultados se difundirán y utilizarán para el mejoramiento de la calidad; estos instrumentos de evaluación serán altamente valorados por maestros, directivos y la sociedad (SEP, 2001: 87).

Los objetivos estratégicos de esta segunda parte del *Programa* incluyen el de *fortalecer la evaluación, el conocimiento y la gestión integral del sistema*, y entre las

políticas correspondientes se incluye una en el sentido de que *se impulsarán decididamente la evaluación y la investigación...* (SEP, 2001: 89-90).

El Instituto Nacional para la Evaluación de la Educación aparece en dos lugares en el apartado de objetivos particulares, líneas de acción y metas, de la Segunda Parte del *Programa*. En relación con el primer objetivo particular, y en concreto como parte de la adecuación de la estructura de la SEP, se indica que *la reorganización de la Dirección General de Evaluación, la creación de la Coordinación General de la Educación Media Superior, y la del Instituto Nacional para la Evaluación de la Educación, se realizarán a más tardar en 2002* (SEP, 2001: 92-93).

Luego, el objetivo particular número cinco de la misma segunda parte establece, entre otras cosas, el propósito de *fortalecer el funcionamiento del Sistema Educativo Nacional, mediante la consolidación del sistema de evaluación*. La primera de las líneas de acción de este objetivo es, precisamente, la de consolidar el Sistema Nacional de Evaluación Educativa, y la primera de sus metas la de *crear el Instituto Nacional para la Evaluación de la Educación, cuyas funciones se articularán con las instancias responsables de la evaluación de la SEP y las entidades federativas* (SEP, 2001: 100).

De manera congruente con los planteamientos anteriores del *Programa Nacional de Educación 2001-2006*, y como resultado de un cuidadoso trabajo de planeación hecho en paralelo, se elaboró el siguiente proyecto para la creación del *Instituto Nacional para la Evaluación de la Educación*, que parte del trabajo de evaluación previamente desarrollado en México y de la mejor la experiencia internacional.

Las primeras versiones del documento se presentaron a especialistas en el tema y luego, a los secretarios de educación y funcionarios equivalentes de las entidades federativas del país, reuniéndose su opinión favorable. También se consideraron las opiniones de los legisladores, el magisterio y el sector privado, en reuniones celebradas para tal cometido.

Por el interés que manifestaron al respecto los señores legisladores, las comisiones de educación del Senado de la República y la Cámara de Diputados, junto con la Secretaría de Educación Pública, organizaron un *Seminario sobre el marco normativo para la calidad educativa*.

Durante el evento, que tuvo lugar los días 23 y 24 de abril de 2002, fue posible escuchar las opiniones

de los mismos legisladores, pero también de representantes del Sindicato Nacional de Trabajadores de la Educación; de las organizaciones nacionales de padres de familia y el sector empresarial, de los organismos de evaluación de la educación superior que funcionan en el país, así como de especialistas mexicanos invitados a propuesta de los diversos partidos representados en el Congreso de la Unión, con lo que pudo garantizarse la pluralidad de los puntos de vista expresados.

La participación de especialistas de organismos internacionales dedicados a la educación y, en especial, a su evaluación, como la UNESCO (a través de su Oficina Regional para América Latina); la Organización de Estados Iberoamericanos (OEI) y la Organización para la Cooperación y el Desarrollo Económicos (OCDE), por conducto de su *Programme for International Student Assessment* (PISA), permitió enriquecer el proyecto con la calificada opinión de los expertos que acudieron en respuesta a la invitación del Congreso mexicano.

En resumen, puede sostenerse que el proyecto del Instituto Nacional para la Evaluación de la Educación ha sido sometido a un proceso de discusión y análisis, por parte de todos los sectores interesados en el tema, poco habitual por su intensidad y amplitud. También puede afirmarse que tal consulta arroja como saldo un consenso prácticamente unánime, en cuanto a la conveniencia de que un organismo como el que se propone se establezca a la brevedad posible; si bien se advierte la necesidad de que cuente con los recursos, el tiempo y la solidez institucional necesarias para que pueda desempeñar en forma adecuada su delicada misión.

En particular se subrayó la necesidad de que el organismo tenga un claro margen de independencia, tanto respecto de las autoridades educativas como de cualquier otra parte interesada, y se coincidió en que es necesario buscar la mejor alternativa para ello dentro del marco jurídico mexicano o, de ser necesario, hacer los ajustes necesarios para que el organismo cuente con el margen de autonomía necesario.

A partir de lo anterior, en las páginas siguientes se presenta el proyecto del Instituto Nacional para la Evaluación de la Educación.

El proyecto comienza recordando las actividades de evaluación hechas por la SEP desde fines de los años



setenta, que son los antecedentes nacionales del INEE y, junto con los internacionales, se recogen en una amplia revisión de literatura que se presenta en el anexo 1, cuyas conclusiones se retoman en el primer apartado siguiente. El segundo punto del documento desarrolla los argumentos que justifican la creación de un organismo con las características del que se propone.

En el punto siguiente se presenta un planteamiento conceptual, en el que se resumen los rasgos del organismo desde un punto de vista técnico, a partir de las experiencias revisadas; después se presentan las características jurídicas y organizacionales del Insti-

tuto, y posteriormente sus perspectivas de desarrollo en tres horizontes: el de su primer año de operación, al final de 2002; el de 2006, al terminar la actual administración federal, y el de seis años después, en 2012. En la conclusión se resumen los principales elementos del proyecto.

El anexo 1 consiste en una amplia revisión de literatura, en la que pueden verse los antecedentes nacionales e internacionales del Instituto, a partir de cuyo conocimiento se plantea el proyecto. El anexo sistematiza una extensa bibliografía especializada sobre el tema de la evaluación.

I. Antecedentes

Desde la década de 1970, la SEP comenzó a realizar actividades que llevaron a la creación de la Dirección General de Evaluación (DGE) y a las primeras evaluaciones del aprendizaje alcanzado por los alumnos de primaria con muestras nacionales. La evaluación, sin embargo, no se consolidó, lo que se manifestó en deficiencias técnicas en los procesos de evaluación, y en la escasa utilización de los resultados.

Hasta principios de los años noventa la experiencia de evaluación incluía el *Estudio Integral de la Evaluación Preescolar, Primaria y Secundaria*, los exámenes de ingreso a educación normal y los de ingreso a secundaria. A fines del sexenio 89-94, la inclusión de los factores de rendimiento escolar y profesional en el esquema de Carrera Magisterial provocó la necesidad de aplicar anualmente pruebas de conocimientos a números considerables de maestros y alumnos.

En el sexenio 1995-2000 la situación empezó a cambiar: algunas entidades comenzaron a desarrollar sistemas de evaluación propios; la SEP, por su parte, se propuso crear un Sistema Nacional de Evaluación Educativa (SNEE) y dieron pasos importantes en esa dirección. Además de continuar las evaluaciones nacionales iniciadas en los setenta y las pruebas de aprendizaje con grandes números de alumnos, en el contexto del programa de Carrera Magisterial, las principales actividades de la DGE en el sexenio 1995-2000 fueron:

- Exámenes de ingreso a secundaria y a educación normal (IDANIS e IDCIENT).
- Aplicaciones de pruebas junto con acercamientos cualitativos, como parte de la evaluación de programas compensatorios (PARE, PAREB).
- Impulso de evaluaciones cualitativas y actividades para desarrollar la capacidad de evaluación de los maestros; impulso a las Áreas Estatales de Evaluación; constitución del Comité Técnico

Interestatal de Evaluación.

- Construcción y aplicación de instrumentos de evaluación del aprendizaje de los alumnos en relación con estándares curriculares nacionales, para primaria y secundaria.
- Con el Ceneval, procesos de selección de alumnos para las instituciones de nivel medio superior de la zona metropolitana de la ciudad de México.
- Participación en proyectos internacionales, en particular en el Tercer Estudio Internacional sobre Matemáticas y Ciencias (TIMSS, de la IEA); las pruebas del Laboratorio Latinoamericano de Evaluación de la Calidad Educativa de la OREALC-UNESCO; y las del *Programme for International Student Assessment* de la OCDE (véase Velázquez, 2000; SEP, 2000).

Los esfuerzos recientes han sido importantes para establecer las bases de una cultura de la evaluación; es claro, sin embargo, que debe aumentar la conciencia de los beneficios que tal actividad puede aportar. Se deberá desarrollar las capacidades técnicas para la medición del logro académico, pero también las que se requieren para la interpretación correcta y el aprovechamiento de los resultados, tanto por parte de las autoridades educativas, como por otros usuarios importantes de la información, en especial los maestros, los padres de familia y los propios alumnos.

Para alcanzar un nivel técnico superior, además de desarrollar la capacidad de las entidades federativas, la estructura actual del SNEE debe mejorar. Es clara la necesidad de un órgano especializado con recursos técnicos y personalidad jurídica los cuales que permitan la realización de todos los aspectos técnicos de la evaluación con la calidad necesaria.

Debe subrayarse la importancia de que la evaluación educativa, además de ser de calidad para ofrecer susten-



to sólido a las decisiones de las autoridades, sirva para que la sociedad en general cuente con información adecuada sobre el funcionamiento del sistema educativo. Para que esta función de rendición de cuentas se cumpla cabalmente los resultados de la evaluación deberán ser del dominio público.

En lo referente a las experiencias internacionales, la revisión del anexo 1 sobre la situación que guardan en la actualidad los sistemas de evaluación educativa en diversos países del mundo; muestra con claridad que la actividad evaluativa es un elemento presente en las políticas educativas de todos ellos.

Aunque está claro que la revisión dista mucho de ser exhaustiva, cubre un número considerable de países y, sobre todo, incluye una buena parte de los más importantes del mundo, tanto en términos cuantitativos, por las dimensiones de su población y su sistema educativo, como cualitativos, por el nivel de desarrollo de dichos sistemas.

La información menos precisa de que se dispone permite añadir que la tendencia señalada abarca también a muchos de los países no incluidos expresamente en esta revisión, así se consideran hay bases suficientes para sostener la aseveración de que la evaluación es un rasgo omnipresente en las políticas educativas de la gran mayoría de los países del mundo.

En contra de algunas opiniones opuestas a las evaluaciones de tipo objetivo en gran escala, lejos de que se estén abandonando, ese tipo de prácticas sigue extendiéndose tanto en el país en donde está más arraigada, los Estados Unidos, como en muchos otros del mundo desarrollado y de los países llamados en vías de desarrollo.

También puede apreciarse que las prácticas evaluativas en modo alguno se limitan al uso de tales instrumentos, sino éstas utilizan diversos acercamientos. En algunos casos las pruebas se usan para evaluar sistemas y subsistemas educativos con propósitos de monitoreo, planeación y rendimiento de cuentas (*accountability*) y en otros para evaluar a los alumnos en lo individual, apoyando a diferentes mecanismos basados en la responsabilidad de los maestros y manejados en cada plantel o distrito escolar.

En este sentido, y como un argumento que complementa las opiniones sobre la evaluación educativa en gran escala, matizando de manera importante las posturas extremas en pro y en contra al respecto, son

ilustrativos los debates que tuvieron lugar en el país con más experiencia al respecto, los Estados Unidos.

En 1999 y 2000 el debate nacional sobre las bondades y peligros de las pruebas que se mantiene vivo desde hace años en ese país, tuvo momentos particularmente álgidos, cuando se comprobó que una de las compañías privadas más importantes en el campo de las pruebas de aprendizaje, había cometido graves errores al reportar los resultados de miles de alumnos en varios estados y distritos escolares, incluyendo al mayor de la nación, estado de Nueva York.

Además de los alumnos, uno de las personas más seriamente afectadas por los errores fue el responsable del sistema escolar de Nueva York, Rudy Crew, entusiasta promotor de las pruebas, quien fue despedido de su cargo. Es interesante conocer su reacción: *la solución es no utilizar el puntaje en una prueba como la única fuente de información sobre el desempeño de un alumno. Éstos son errores humanos, y volverán a ocurrir* (Steinberg y Henriques, 2001).

Pese a las graves consecuencias que los errores de las pruebas tuvieron para él, no consideró que éstas debían suprimirse; simplemente señaló que nunca deben utilizarse de manera única para tomar decisiones de gran impacto, como siempre han recomendado las mejores prácticas de evaluación.

Una lección adicional que se puede sacar es que la complejidad técnica de las evaluaciones hace que incluso organismos fuertes puedan tener fallas, sobre todo si se manejan con criterio comercial y lucrativo. Por ello, la existencia de un ente público de alto nivel técnico, será un elemento regulador importante del mercado de pruebas que como dejan ver las tendencias internacionales, probablemente comenzará pronto a surgir en el país, sobre todo si el Estado mexicano no toma la iniciativa en este sentido.

También contra opiniones frecuentes en el medio, la revisión de la experiencia internacional muestra con claridad que la evaluación no debe asociarse necesariamente con las tendencias hacia la privatización de la educación pública; por el contrario, se trata de un rasgo que comparten tanto los países que se han orientado en esa dirección, como los que siguen apoyando la existencia de sistemas de educación pública fuertes. Es claro en este sentido el caso de muchos países europeos, especialmente los escandinavos, que ocupan luga-

res destacados en las evaluaciones internacionales de la calidad educativa.

La revisión indica también que el estatus jurídico y la ubicación organizacional de la unidad responsable de la evaluación puede ser muy variada. Como se aprecia en el cuadro uno, lo más usual es que sean oficinas situadas en el seno de los respectivos ministerios quienes se encarguen de la evaluación, si bien aparecen dos tendencias:

- Por una parte, surgen entidades públicas que, o siguen siendo parte del ministerio; o son organismos separados con cierta dependencia con respecto a las autoridades educativas (órganos descentralizados o de estatus similar), pero tienen considerable independencia técnica. Es el caso del INCE español, del CITO holandés, de la ACER australiana y del INEP brasileño. Dada

la peculiar situación de la autoridad educativa en el Uruguay (agencia pública que no depende del titular del ejecutivo), el estatus de la oficina encargada de la evaluación es también de mayor independencia en relación con la que tiene normalmente una dependencia del ministerio.

- En otros casos, algunas tareas especializadas, sobre todo las de mayor complejidad técnica, como la elaboración de instrumentos de evaluación de alta calidad, se encomiendan a entidades diferentes completamente independientes del ministerio, como el ETS americano, empresa no lucrativa que maneja el *National Assessment of Education Programme* de los Estados Unidos, o la agencia elaboradora de pruebas de la Universidad de Cambridge, que apoya al Ministerio de Educación de Singapur en ese sentido.

CUADRO I. RASGOS ESENCIALES DE ORGANISMOS DE EVALUACIÓN

PAÍS	ÓRGANO EVALUADOR	ESTATUS JURÍDICO	COBERTURA				
			Niveles educativos	Áreas curriculares	Grados	Población	Periodicidad
Australia	ACER	Empresa	Básica	L-M-CN-CS	Clave	Muestra	
España	INCE	OPD	Básica	L-M-CN-CS	Clave	Muestra	Multianual
Francia	DPDP y M des Ex.	Ministerio	Básica media	Todas	Clave	M y C	Anual
Holanda	CITO	OPD	Básica	Todas	Clave	M y C	Multianual
Singapur	Ministerio	Minist-Univ	Básica	Todas	Clave	Censo	Anual
USA	NAEP-ETS	Empresa	Básica	L-M-CN-CS	Clave	M (2005 C)	Multia (2005 a.)
Argentina	Ministerio	Ministerio	Básica	L-M-CN-CS	Clave	M y C	Anual
Bolivia	SIMECAL	Ministerio	Básica	L-M	Clave	Muestra	Anual
Brasil	INEP	OPD	Todos	L-M-CN-CS	Clave	Muestra	Multianual
Chile	SINCE	Minist-Univ	Básica	L-M-CN-CS	Clave	M y C	Anual
Colombia	ICFES	OPD	Todos	L-M-CN-CS	Clave	Muestra	Anual
Costa Rica	IIMEP-MEP	Minist-Univ	Básica	L-M-CN-CS	Clave	M y C	Anual
Ecuador	APRENDO	Ministerio	Básica	L-M	Clave	Muestra	Anual
Honduras	UPN	Minist-Univ	Básica	L-M	Clave	Muestra	Anual
Perú	Ministerio	Ministerio	Básica	L-M-CN-CS	Clave	Muestra	Anual
Uruguay	UMRE	OPD	Básica	L-M-CN-CS	Clave	Censo	Multianual

Abreviaturas:

OPD = Organismo público descentralizado, no autónomo.

Áreas curriculares: L = Lengua; M = Matemáticas.

CN = Ciencias Naturales; CS = Ciencias Sociales.

Grados: Clave = finales de niveles educativos: primaria, secundaria, media.

Población: M = Muestra; C = Censo.



El cuadro 1 muestra un alto grado de coincidencia en varios aspectos técnicos:

- Con excepción de Brasil y Colombia, la evaluación de la educación superior se maneja en organismos diferentes a la educación básica, y aún en esos lugares hay considerable independencia entre ambos tipos de evaluaciones, y una extendida opinión en el sentido de que parece más conveniente separarlas.
- En todos los casos se evalúan solamente algunos grados del sistema educativo, los que se consideran clave, por lo general los últimos de niveles o ciclos. El número de grados evaluados varían desde dos hasta cinco, pero en ningún caso se evalúan mediante estos sistemas, todos los grados.
- También hay coincidencia al evaluar las áreas consideradas claves del currículo, en particular las de Lengua, Matemáticas y Ciencias, si bien hay diferencias en cuanto al contenido de la última (no siempre se incluyen las mismas Ciencias Sociales). Se estudian en diversos casos otras áreas, y en algunos sistemas pequeños y/o muy fuertes se evalúan todas, como en Singapur, Holanda y, de otra manera, en Francia.
- En esos casos también la evaluación es anual y censal, puesto que tiene carácter sumativo, pero sólo en referencia a pocos grados clave: el último de la primaria, la secundaria y la enseñanza media.
- En la mayoría de los sistemas las evaluaciones se aplican a muestras y en ciclos multianuales, lo que es congruente con su naturaleza formativa o diagnóstica. Destaca el caso de los Estados Unidos, en donde hasta lo actual el NAEP evaluó muestras nacionales de alumnos en grados y materias clave, en ciclos multianuales; a partir del año 2005, según la nueva Ley Bush, el NAEP seguirá haciendo evaluaciones muestrales, pero los estados deberán participar obligatoriamente en ellas, y además deberán hacer sus propias evaluaciones censales y anuales en 4°, 5°, 6°, 7° y 8° de la educación básica.
- Los sistemas de evaluación comenzaron a funcionar con menos áreas y grados, y en forma muestral y multianual; paulatinamente se añadieron áreas curriculares y grados, y en algunos

casos se tiende a la evaluación censal y anual de grados clave.

Es fácil apreciar que los rasgos básicos del organismo de evaluación que se pretende establecer en México, coinciden con los que la experiencia internacional muestra como más adecuados para este tipo de entidades.

Volviendo al punto de la independencia del organismo, puede afirmarse que lo esencial no es el estatus jurídico que tenga, pero sí su capacidad técnica y su solvencia moral, de las cuales dependerá su credibilidad.

De Landsheere, citando a Guthrie, considera que lo ideal es que las actividades evaluativas sean desarrolladas por un organismo no directamente dependiente del Ministerio (apolítico), no vinculado con grupos de presión, institucionalmente creíble, al tanto de la problemática educativa, metodológicamente competente y que tenga capacidad de sintetizar los resultados de su trabajo, y de difundirlos entre los diversos grupos interesados: investigadores, especialistas en evaluación, maestros, funcionarios, padres de familia, medios de comunicación y público en general (1994: 43).

Después de comparar los casos de España, Francia, Inglaterra, Suecia y Portugal, otro especialista encuentra que en todos los casos se han establecido instituciones que *dependen de los más altos niveles de los ministerios y gozan, en general, de una gran independencia científica y metodológica*, y señala la necesidad de cualquier organismo de esta naturaleza por cumplir con dos principios básicos: *que tenga una elevada independencia política a la hora de diseñar un modelo de evaluación del sistema educativo y de llevarlo a término... tanto en su práctica diaria, como en la percepción que la sociedad tiene de este organismo; y un elevado nivel técnico-pedagógico, para garantizar la calidad de los procesos y la validez, fiabilidad y credibilidad de sus resultados*. (Mateo, 2000: 244)

El cúmulo de experiencias revisado, que se resume en las páginas anteriores, contiene sin duda referentes suficientes para sustentar el diseño de un organismo de evaluación para México, que tenga las características necesarias para asegurar la calidad de su trabajo y, por ende, maximizar las posibilidades de que tenga un efecto positivo en la educación nacional.



Esa revisión conlleva a la conclusión de que no hay fundamento para la visión, frecuente en nuestro medio, de la evaluación como una tarea corta de miras, limitada al uso de pruebas de papel y lápiz, con preguntas de opción múltiple sobre aspectos superficiales del currículo, que dejarían fuera las dimensiones educativas más importantes y, por tanto, tendrían consecuencias nocivas para una educación concebida en una forma más rica.

Por el contrario, la historia de la evaluación en el terreno internacional muestra que se trata de un campo en constante avance, y ha conocido desarrollos teóricos y técnicos fundamentales en la segunda mitad del siglo XX; mediante su utilización la evaluación puede ser un elemento de gran valor para el mejoramiento de las escuelas en todas las dimensiones del quehacer educativo, incluyendo las más finas y profundas.

Las críticas que suelen hacerse contra la evaluación en general, y contra las pruebas de tipo objetivo en particular, no carecen de fundamento, pues no es raro que se manejen en forma deficiente, dando lugar a interpretaciones inadecuadas de los resultados obtenidos, que pueden tener serias consecuencias para las escuelas y los alumnos, como ocurrió en los Estados Unidos, según se comentó antes.

En un nivel más profundo, es importante entender que la complejidad de los procesos educativos no puede ser captada en forma adecuada por un solo acercamiento, por lo que es necesario emplear varios, y combinarlos prudentemente para obtener una visión lo más completa posible, y ofrezca fundamento sólido para las decisiones y acciones de mejoramiento.

Pero concluir, a partir de las deficiencias reales o supuestas de las pruebas, que no deben de utilizarse para la evaluación de sistemas de educación básica tan grandes como el de México –el séptimo del mundo, con más de veinte millones de alumnos– es, sin duda, un error grave, que contradice la experiencia internacional más sólida.

No es aceptable la falsa disyuntiva que contraponen pruebas en gran escala pobremente diseñadas a evaluaciones finas en el nivel de aula o escuela, como si no pudiera haber otras opciones.

En cambio, es posible construir y aplicar pruebas objetivas para aplicaciones en gran escala de excelente calidad, así como interpretarlas y emplearlas correctamente. Ello requiere, sin duda, de un esfuerzo técnico considerable, pero su utilización en un país como México será un elemento de gran valor para el mejoramiento del sistema educativo.



2. Justificación

La argumentación en que se sustenta el proyecto puede resumirse de la manera siguiente:

Sobre la necesidad de un buen sistema nacional de evaluación educativa:

- a) Para que exista un desarrollo nacional relevantes, tanto en lo económico como en lo político y lo social, es necesario un sistema educativo de elevada calidad.

Ahora bien, un buen sistema de evaluación, que posibilite la recolección, el análisis y la interpretación de información confiable, válida y relevante; el cual detecte avances, aciertos, dificultades y disfuncionalidades del proceso educativo, y que, por lo tanto, constituya una base sólida para la toma de decisiones estratégicas adecuadas para sustentar un avance progresivo en el logro de los objetivos educativos, es un requisito necesario de una educación de calidad.

Por lo tanto, la existencia de un buen sistema de evaluación para la educación básica debe ser una prioridad de las políticas públicas de México en el área educativa.

- b) Sobre la necesidad de que se disponga de recursos adecuados para la evaluación:

Para que sus resultados sean útiles, la evaluación educativa se debe hacer en forma permanente y con alto nivel de calidad.

Ahora bien, desarrollar instrumentos de evaluación acordes con estándares de calidad aceptados internacionalmente; aplicarlos, analizar sus resultados y hacer investigación sobre ellos, son tareas de alta complejidad, requieren gran tiempo y suponen una gran capacidad técnica.

Por consiguiente, la evaluación de un sistema de educación básica tan grande como el mexicano implica la existencia de personal especializado, que cuente con los recursos necesarios y dispon-

ga de la estabilidad y continuidad que el trabajo exige.

- c) Sobre la necesidad de que se constituya un nuevo organismo independiente:

Las consecuencias positivas o negativas de la evaluación educativa pueden ser serias, por lo que los actores implicados, incluyendo a la autoridad federal y las estatales, así como los maestros, constituyen partes con explicable interés en los resultados.

Ahora bien, si las autoridades son a la vez las responsables de la evaluación, ésta no tendrá independencia para gozar de credibilidad frente a otras partes interesadas, como los legisladores, los padres de familia, la sociedad en general y las instancias internacionales.

Por consiguiente, para que el sistema de evaluación apoye de manera efectiva la rendición de cuentas a la sociedad por parte de las autoridades educativas, se requiere también, sea manejado por una entidad que goce de independencia, y así sus resultados sean creíbles; deberá precisarse, como se hará más adelante, el rol del nuevo organismo frente al papel de las autoridades educativas federal y estatales, responsables de las decisiones, incluyendo las de la evaluación, en este caso con apoyo del INEE.

En conclusión, la propuesta para crear un nuevo organismo de evaluación, que funcione con independencia técnica, se desprende de los razonamientos anteriores:

Por una parte, los requerimientos que implica la complejidad de la tarea -en particular los relativos a la continuidad de los proyectos y la estabilidad del personal de alto nivel- se atenderán mejor en una entidad que se maneje con base en criterios académicos.

Por otra parte, en lo que se refiere a la credibilidad y la rendición de cuentas, es clara la ventaja de



un organismo cuyo trabajo técnico, sin estar supeditado a la SEP, apoye su labor y ponga a su alcance, y el de otras partes interesadas, la información necesaria para valorar de manera objetiva el nivel alcanzado por el Sistema Educativo Nacional y los subsistemas, estatales entre otros.

Para tener un buen sistema de evaluación se requiere un organismo con características técnicas que

aprovechen los antecedentes nacionales en el campo y la mejor experiencia internacional al respecto, para que alcance, en un plazo razonable, los más exigentes estándares de calidad, con el estatus jurídico y las condiciones estructurales necesarias para asegurar su autonomía.

Sólo de esa forma estará a la altura de los retos que deberá enfrentar.



3. Planteamiento conceptual

3.1 La calidad educativa y sus dimensiones

El diseño de un sistema de evaluación debe comenzar con una definición precisa del concepto *calidad*, evitando los acercamientos reduccionistas y simplistas que fácilmente se ponen de moda.

En efecto, evaluar la calidad de algo tan importante como la educación nacional, requiere de indicadores que reflejen adecuadamente la medida en la cual el quehacer educativo se acerca a las metas, y a los objetivos trazados en términos de los aprendizajes deseables. La utilización de tasas brutas y cohortes aparentes para las estimaciones de cobertura y eficiencia terminal pueden ser útiles en determinados contextos, pero resultan insuficientes para medir la calidad de la educación y el logro académico de los estudiantes.

De igual forma, el recuento de insumos (como el número de computadoras existentes en las escuelas) resulta insuficiente para evaluar la gestión del sistema. Por el contrario, deberá quedar claro, para captar la compleja y elusiva realidad que pretende expresar la noción de calidad educativa, ésta deberá comprender varias dimensiones que la experiencia reconoce y la literatura expresa de diversa manera.

En forma sintética, un sistema educativo de calidad es el que:

- Establece un currículo adecuado a las necesidades de la sociedad en donde se ubica, entendiendo que deben incluirse las de una mayor productividad del aparato económico, pero también otras fundamentales, en una perspectiva de desarrollo integral y humanista, como las concernientes con la democracia política, el respeto de los derechos humanos, el desarrollo de la ciencia, el cuidado del medio ambiente y la preservación y enriquecimiento de la diversidad cultural. (*Relevancia*)

- Logra que la más alta proporción posible de destinatarios –definidos en términos de grupos por edad o de otra forma– acceda a la escuela, permanezca en ella hasta el final del trayecto previsto y egrese alcanzando los objetivos de aprendizaje establecidos. (*Eficacia interna y externa*)
- Consigue que los aprendizajes logrados por los alumnos sean asimilados por éstos en forma duradera y deriven en comportamientos sociales sustentados en los valores de libertad, equidad, solidaridad, tolerancia y respecto a las personas, que son fructíferos para la sociedad y al propio individuo, quien podrá así alcanzar un desarrollo pleno en los diversos roles que habrá de desempeñar como trabajador, productor, consumidor, padre de familia, elector, servidor público, lector y telespectador, entre otros; en pocas palabras, como ciudadano cabal. (*Impacto*)
- Cuenta, para lograr lo anterior, con recursos humanos y materiales suficientes, y los aprovecha de la mejor manera posible, evitando despilfarros y derroches. (*Eficiencia*)
- Tiene en cuenta la desigual situación de alumnos y familias, de las comunidades en donde viven y las escuelas mismas, y ofrece apoyos especiales a quienes lo requieren, para que los objetivos educativos sean alcanzados por el mayor número posible. (*Equidad*)

En términos técnicos, la educación será de calidad si es *relevante*, si tiene *eficacia interna y externa*, si tiene un *impacto* positivo en el largo plazo, si es *eficiente* en el uso de los recursos y si busca la *equidad*.

Debe añadirse que el concepto de calidad es *relativo* y *dinámico*:

- *Relativo*, porque el juicio sobre la calidad de algo depende del punto de referencia que se tome; por ejemplo un país más desarrollado, uno de un



nivel de desarrollo similar, la situación del propio sistema en un momento dado del pasado, o determinadas metas para un futuro prefijado del propio sistema;

- *Dinámico*, porque nunca se alcanza la calidad absoluta: siempre es posible proponer metas más elevadas y cuando se alcanza una hay razón para desear ir adelante. El punto de referencia más pertinente para valorar la calidad educativa es, justamente, el pasado y el futuro del propio sistema: uno de calidad es el que mejora siempre con respecto a sí mismo.

Conceptualizada así, la calidad no es un estado sino una tendencia: es una *autoexigencia permanente y razonable de superación*, que no se puede adquirir del exterior, sino, sólo puede surgir del interior mismo del sistema a mejorar. La calidad del sistema nacional de educación se alcanzará en cada escuela y aula, o no se alcanzará.

Y para que sea posible ese proceso de superación constante es indispensable contar con un buen sistema de evaluación, de carácter permanente también; gracias al cual las autoridades y los actores del sistema escolar, vigilen regularmente el estado en que se encuentra la calidad educativa y puedan monitorearla permanentemente, para actuar en consecuencia.

3.2 Las variables que influyen en la calidad educativa

La evaluación educativa no constituye un fin en sí misma; es un medio —muy importante, pero sólo un medio— para el mejoramiento de la enseñanza. Ahora bien, para propósitos de mejoramiento la descripción de la situación de un sistema educativo no es suficiente; se necesitan acercamientos *explicativos*, estudios que se acerquen a las *causas* de la situación descrita.

La detección de los *factores* del rendimiento escolar, la valoración del peso de cada uno, y el conocimiento de su *modus operandi* y de la forma cómo interactúa con los demás, permitirán que los tomadores de decisiones cuenten con bases sólidas para diseñar estrategias de mejoramiento adecuadas.

La investigación educativa clasifica de diversos modos esos factores: a fines del siglo XIX y principios del XX, se privilegiaban los factores heredita-

rios o genéticos, en contraposición a los ambientales; el optimismo de los años que siguieron a la Segunda Guerra Mundial hizo albergar altas expectativas en relación con la posibilidad de la educación para promover la movilidad y el progreso social, haciendo que el peso más importante de los factores del rendimiento se asignara a la escuela.

Tras la expectativa que representó el lanzamiento del *Sputnik*, en 1957, vino en 1966 el impacto del informe Coleman, el cual más allá de sus cualidades, por limitaciones metodológicas las cuales se identificaron con precisión posteriormente parecía mostrar que los factores de la escuela no influían prácticamente en el aprendizaje de los alumnos.

Una ola de escepticismo sobre el potencial transformador de la escuela frente a las desigualdades sociales se extendió, resurgiendo las viejas ideas geneticistas, en polémica con posturas radicales, según las cuales únicamente cambios sociales de alcances revolucionarios podrían transformar las posibilidades educativas de los niños de hogares desfavorecidos.

Como resultado de la polémica se desarrolló un creciente interés por estudiar con mayor detalle y profundidad, los factores de la escuela, sin limitarse a los más visibles, como los insumos materiales o la escolaridad de los maestros, sino incluyendo dimensiones y variables más finas, como las prácticas docentes, las actitudes y expectativas de los diversos actores educativos, y hasta los procesos que tienen lugar en el interior de la mente, con base en los acercamientos más avanzados de la Psicología y la Pedagogía, que escudriñan los procesos metacognitivos de alumnos y maestros.

Los desarrollos de las últimas décadas llevaron a posturas más equilibradas, que reconocen la importancia, tanto de los factores llamados *de la demanda educativa* —o sea los que corresponden a los alumnos, sus familias y el medio social en que se sitúan— así también a los designados como factores *de la oferta educativa*; o sea, los que se refieren a la escuela, incluyendo currículo, materiales, maestros, prácticas pedagógicas, etcétera.

En el terreno metodológico, el interés por aspectos de la realidad más difíciles de observar trajo consigo, desde los sesenta, el desarrollo de acercamientos cualitativos a la investigación que, tras las respectivas polémicas con los acercamientos llamados cuantita-



tivos, está dejando el lugar a toda una gama de acercamientos mixtos.

El estudio de los factores con influencia en la calidad educativa puede hacerse, por cierto, en el marco de investigaciones los cuales no sean de carácter evaluativo. Es razonable también, con todo, que los trabajos de evaluación incluyan estudios sobre variables clave, tanto del lado de la *demand educativa* como, sobre todo, de la *oferta*, para complementar la perspectiva descriptiva de la evaluación con acercamientos explicativos.

Además, debe señalarse, que los desarrollos de la Psicología y, más ampliamente, de las Ciencias cognitivas y la Informática, están dando lugar a nuevas concepciones de la inteligencia y la mente que tendrán profunda influencia, a su vez, en las concepciones de la educación y el aprendizaje y, por supuesto, en lo relativo a su evaluación.

3.3 El universo a cubrir

Además de las dimensiones del concepto de calidad y las variables explicativas de la misma, el otro elemento fundamental para la definición de los rasgos centrales de un sistema de evaluación está constituido por las dimensiones y variantes del objeto a evaluar, en este caso el Sistema Educativo Mexicano.

Para precisar el propósito del INEE, en un contexto de recursos limitados como México, es importante evitar la duplicidad de esfuerzos y recursos, y las posibles contradicciones con otros actores en el ámbito de la evaluación, por lo cual se requiere considerar los avances desarrollados en este terreno, y las instituciones, programas y proyectos que operan con éxito en el país.

En la última década han surgido en México varios organismos especializados los cuales tienen que ver con la evaluación de la educación superior, incluyendo al Ceneval, los CIEES, los mecanismos propios de varios subsistemas, como los de la ANUIES, el COSNET y la FIMPES, y diversas agencias acreditadoras. De acuerdo con las tendencias internacionales al respecto, la SEP y la ANUIES, a través de la CONPES, dieron un paso importante hacia la integración de esos organismos en un sistema nacional de evaluación del nivel educativo más elevado, con la constitución del Consejo para la Acredi-

tación de la Educación Superior (Copaes) a finales del año 2000.

Se puede concluir, que en este nivel educativo ya existen actores y procesos con experiencia y capacidad para la evaluación y que, por lo tanto, y en forma congruente, con la experiencia internacional, no será necesario que el INEE tenga entre sus objetivos la evaluación de la educación superior.

En lo concerniente a la educación media superior, en ese nivel educativo existen aproximaciones valiosas, pero que se limitan a la aplicación de pruebas estandarizadas de ingreso y egreso (EXANI I y EXANI II del Ceneval). En cambio, se carece de estudios sobre la calidad de las escuelas, los procesos educativos y el subsistema como tal. Por ello el *Programa Nacional de Educación 2001-2006* contempla la creación de un organismo especializado para la evaluación de planteles y procesos en ese nivel, pero no en lo que hace a la elaboración de pruebas de rendimiento. El Ceneval podrá seguir cubriendo parcialmente ese campo con los EXANI I y EXANI II.

En el campo de la evaluación de la educación básica, los logros hasta el presente por la SEP, resultan insuficientes, sobre todo en lo referente a la existencia de mecanismos que gocen de credibilidad por su independencia. Por ello parece clara la necesidad de nuevos elementos, como los contemplados como tarea del INEE.

Por lo dicho, se contempla que el ámbito de acción del nuevo organismo no incluya la educación superior, sino la básica, con los niveles de preescolar, primaria y secundaria. En cuanto a la media superior, no se descarta que el INEE la atienda en el futuro, evitando duplicidades con el Ceneval, pero al menos inicialmente su atención deberá limitarse a la enseñanza básica. Teniendo en cuenta la tendencia a la educación obligatoria, en su extensión hasta los 18 años, es previsible que la media superior llegue, tarde o temprano, a definirse también como básica; en cuyo caso, caería directamente dentro del ámbito de competencia del INEE; aún antes de que tal ocurra, el INEE podría cubrir parcialmente dicho nivel con sus pruebas, una vez que atienda suficientemente los niveles anteriores.

En referencia a la educación básica, para crear un sistema integral, los mecanismos de evaluación del INEE deberán ser diseñados de manera que tengan



en cuenta todos los aspectos y elementos de dichos niveles educativos, si bien, por volumen y la complejidad de los mismos, lo anterior no podrá hacerse evaluando cada año a todos los alumnos de preescolar, primaria y secundaria de México, en todas las materias o áreas del currículo.

Como se ha apuntado y puede verse con amplitud en el anexo 1, ni siquiera los sistemas de evaluación más fuertes proceden de esa manera, sino que seleccionan grados especialmente importantes (como los últimos de la primaria y la secundaria) y áreas clave del currículo (por ejemplo: Matemáticas y Lectoescritura), para desarrollar instrumentos que se aplican a muestras nacionales cuidadosamente diseñadas para asegurar su representatividad, teniendo en cuenta las diferencias regionales, étnicas entre otras.

Para cubrir todas las áreas del currículo deberá manejarse el concepto de *ciclos*, de manera que en un año se apliquen pruebas más amplias de una materia o área, junto con instrumentos de menor extensión de otras; al año siguiente se pondrá el énfasis en otra materia, y así sucesivamente. Ciclos semejantes pueden diseñarse para evaluar en forma periódica muestras de alumnos de los diversos grados y, también con la periodicidad que se juzgue conveniente, inclusive a todos los alumnos de grados claves de ciertos subsistemas estatales, incluso a todos los de algún grado del Sistema Educativo Nacional.

Así deberá proceder el INEE, para hacer un trabajo de la calidad deseada; de manera progresiva podrán cubrirse áreas más difíciles de evaluar, como la enseñanza artística. A más de la imposibilidad de evaluar anualmente, con alta calidad, a los más de veinte millones de alumnos del sistema nacional de educación básica, en relación con todos los aspectos del currículo, la evaluación cíclica de áreas y grados se justifica si se tiene en cuenta que el propósito del Instituto no será el de evaluar individualmente a los alumnos sino, muy precisamente, *al Sistema Educativo Nacional y los subsistemas estatales y otros*.

Los aspectos del sistema de educación básica que el INEE deberá tener en cuenta para su planeación a mediano y largo plazos son:

- Todos los grados de preescolar, primaria y secundaria; luego de media superior.
- Todas las áreas o asignaturas del currículo: Matemáticas, Español, Ciencias, etcétera.

- Los aspectos cognoscitivos, pero también los afectivos y psicomotores, incluyendo todos los niveles de complejidad, con especial atención a los más elevados, de manera que se privilegie no la memorización o acumulación de información, sino la capacidad de localizarla, procesarla y sintetizarla, y la de aplicarla en la práctica.
- Todas las entidades de la República, con una representación adecuada de su variedad, incluyendo zonas urbanas y rurales, del norte, centro o sur, comunidades indígenas, etcétera.
- Todas las modalidades del servicio: presencial o a distancia, escuelas públicas y privadas, matutinas y vespertinas, de organización completa e incompleta, etcétera.

Como implica el concepto de calidad propuesto, el sistema de evaluación del INEE no se limita a la aplicación de pruebas, sino que incluye la obtención de información sobre insumos del sistema, sobre aprendizajes no cognitivos de los alumnos, sobre procesos, e incluso sobre necesidades del entorno, para poder construir indicadores de relevancia, eficacia interna y externa, impacto, eficiencia y equidad.

Por consiguiente en lo relativo a lo faltante a cubrir no debe entenderse únicamente en cuanto a las pruebas, sino también en cuanto a los otros dos componentes estructurales del INEE –indicadores y evaluaciones de escuelas– como se desarrolla a continuación.

3.4 Los componentes del INEE

Las funciones sustantivas del INEE se desarrollarán en tres subsistemas básicos: el de indicadores de calidad, el de elaboración y aplicación de pruebas de aprendizaje y el de evaluación de escuelas. El subsistema de indicadores es el componente que dará unidad al conjunto, si bien son los otros dos –el subsistema de evaluación de escuelas y, sobre todo, el de pruebas de aprendizaje– los que constituirán la mayor parte del trabajo del Instituto. A continuación se explica la relación entre los tres subsistemas, y posteriormente se describe cada uno.

Para evaluar un sistema educativo, atendiendo las dimensiones del concepto de calidad presentado antes, debe contarse con información sobre muchos elementos, la cual debe relacionarse.

Para conjuntar tales características los indicadores no pueden reducirse a un listado reducido, mal estructurado y con información de dudosa calidad, sino que deberán formar un conjunto muy bien integrado, que se derive lógicamente de un concepto de calidad rico y preciso, y que se caracterice por la elevada calidad de los procesos de recopilación, integración, análisis y presentación de la información.



La necesidad de que los indicadores sean de fácil interpretación no debe impedir la atención de su consistencia; serán necesarios, además, esfuerzos específicos para encontrar formas de difusión claras e inteligibles, y para desarrollar la capacidad de interpretación de los usuarios.

Los sistemas de indicadores suelen basarse en información documental recabada por razones administrativas –como el número de alumnos a inicio y fin de un ciclo escolar, obtenido mediante las formas 911– que es insuficiente para construir los indicadores más importantes, los cuales se refieren a los resultados del aprendizaje.

Por ello un buen sistema de indicadores necesita de los otros dos subsistemas que contempla el INEE, el de pruebas y el de evaluación de escuelas, pues sólo mediante ellos podrá tener, además de los insumos, elementos relativos a los productos y los procesos del sistema escolar, especialmente en sus aspectos cualitativos, que la documentación administrativa no recoge de manera confiable, como el nivel de logro de los objetivos curriculares por parte de los alumnos, o el grado de esfuerzo y compromiso de los maestros que reflejan las prácticas docentes.

En el INEE se concebirán los subsistemas de pruebas y escuelas, los cuales podrían funcionar de manera autónoma, con su lógica y tiempos propios, como elementos que se integrarán en el conjunto, aportando elementos precisos. A continuación se describen brevemente los tres subsistemas, sobre los que podrán verse mayores detalles en el anexo uno, que recoge la experiencia internacional en relación con los tres tipos de componentes de un sistema integral de evaluación.

El subsistema de indicadores

Según Mateo, *los indicadores son datos..., recogidos de manera sistemática, que informan respecto de los recursos impartidos, los procesos realizados o las metas conseguidas, en relación con los objetivos de calidad que se había propuesto una organización, en este caso un centro educativo* (2000: 169).

En el caso del Sistema Nacional de Evaluación Educativa, un subsistema de este tipo tendrá como objetivo proporcionar información sobre el estado y la evolución del sistema educativo, en la forma de un conjunto de indicadores que cubra todas las dimen-

siones del concepto de calidad y todos los niveles y componentes del sistema, presentándolos de manera sintética y continuada, de manera que puedan hacerse análisis longitudinales, y en forma congruente con los sistemas internacionales, para así realizar comparaciones transversales con otros países y sistemas.

Con tal fin hay información que recogen o elaboran instancias preexistentes, como la Dirección General de Planeación, Programación y Presupuesto de la SEP (DGPPP), y sus contrapartes en las entidades federativas, así como el Instituto Nacional de Estadística, Geografía e Informática (INEGI), el Consejo Nacional de Población (Conapo), etcétera.

Con el trabajo de los dos subsistemas siguientes, el INEE podrá aportar, a su vez, otros elementos fundamentales sobre resultados de aprendizaje y sobre procesos que tienen lugar en las escuelas, para la construcción de indicadores.

Como establece el *Programa Nacional*, la responsabilidad de diseñar y operar el Sistema Nacional de Indicadores Educativos será compartida por las instancias respectivas de la SEP y el INEE, y deberá articularse con el Sistema Nacional de Indicadores de la Presidencia de la República (SEP, 2001: 101-102).

La SEP, por conducto de la Dirección General de Planeación, Programación y Presupuesto, de la Subsecretaría de Planeación y Coordinación, maneja tanto el sistema de información como el de indicadores, éstos pueden consultarse en diversas formas, incluso a través de la página electrónica de la SEP (www.sep.gob.mx). Ambos sistemas deben, sin embargo, perfeccionarse, como plantea el *Programa Nacional*, actividad en la que deberá participar el INEE.

Además de aportar los resultados de las pruebas que aplique, el INEE participará en el diseño de nuevos indicadores y el mejoramiento de los existentes, a partir de la información disponible, y podrá sugerir formas de mejorar la recolección de la misma; en caso de que no exista información para la construcción de ciertos indicadores deberán definirse los medios para generarla. Deberán también monitorearse los procesos actuales de obtención de información para asegurar su confiabilidad.

En el marco de respeto al papel de la SEP, y en una perspectiva de apoyo, el trabajo del INEE relativo a indicadores no substituirá al de la DGPPP, éste seguirá manejando indicadores para la planeación y



control del Sistema Educativo Nacional, sino que lo enriquecerá, aportando el complemento de una perspectiva con algunas dimensiones de la calidad que los indicadores habituales no incluyen. Por su carácter de organismo técnicamente independiente, el INEE será también un elemento esencial del Sistema Nacional de Indicadores Educativos, puesto que podrá aumentar su credibilidad.

El subsistema de pruebas

Este subsistema del INEE tendrá como objetivo elaborar instrumentos de evaluación del aprendizaje de conocimientos, habilidades, actitudes y valores de los alumnos en todas las áreas del currículo, para aplicarlos y analizar e interpretar los resultados obtenidos. Elaborar pruebas que satisfagan los estándares de calidad los cuales permiten los avances de la psicometría es una tarea que requiere de gran capacidad técnica y logística, si se piensa que, a mediano plazo, se requerirán instrumentos variados:

- De aptitudes (Lectura, Razonamiento verbal y matemático) y de conocimientos.
- Que cubran las diversas áreas del currículo (Lectura, Expresión escrita, Comunicación oral, Matemáticas, Ciencias Naturales, Historia, Geografía y Ciencias Sociales, Actitudes cívicas y valores, Educación artística, Educación física, etcétera).
- Consideren toda la gama de niveles de dominio, y no sólo los más elementales.
- Que cubran todos los grados de la educación básica, aplicándose a millares de alumnos de todas las regiones del país, que representen al universo, en forma casi simultánea.
- Y su aplicación se repita anualmente, en ciclos por definir.
- Que puedan aplicarse (las mismas pruebas u otras *ad hoc*) a adultos (para evaluar, por ejemplo, su nivel de lectoescritura (alfabetización funcional), o sus conocimientos de Matemáticas u otras áreas, como alfabetización aritmética, científica, computacional, en inglés, etcétera.) ya que la educación destinada a ellos es parte del sistema educativo.

Como se mencionó, el trabajo del Instituto no podrá cubrir todos los aspectos en forma completa

cada año. La complejidad técnica del asunto hace que ni siquiera los países más ricos tengan una cobertura de ese tipo, por lo cual deberán definirse periodos multianuales a lo largo de los cuales se vayan cubriendo los diversos aspectos (áreas del currículo y grados especialmente) en forma planificada, en forma cíclica. Es precisamente la complejidad técnica la que hace necesario un organismo especializado para la elaboración de las pruebas, y las aplique a muestras o poblaciones del país, y las entidades, mediante convenios con las autoridades educativas nacionales y estatales; asimismo procese los resultados, cuidando todos los aspectos de estandarización y construcción de variantes equivalentes, para permitir la comparación de los resultados del sistema nacional y los subsistemas estatales, en forma transversal y a lo largo del tiempo.

En este punto conviene aclarar una duda que se expresa reiteradamente en relación con las pruebas de aprendizaje elaboradas y aplicables por el INEE. El indicar que la tarea del *nuevo organismo* no será la evaluación individual de alumnos, maestros o escuelas, sino la del Sistema Educativo Nacional, o los sistemas estatales como tales, lleva a algunos a preguntarse cómo será posible evaluar el sistema si no se evalúa a sus integrantes.

Por esto conviene precisar que, obviamente, deberá evaluarse cierto número de alumnos, de determinadas escuelas, atendidos por maestros concretos; sin embargo, el hecho de que, por regla general, dadas las dimensiones del sistema, se deba trabajar con muestras y en ciclos multianuales, hace que las evaluaciones individuales de alumnos, maestros y escuelas no puedan hacerse con base en las evaluaciones del INEE.

Aun si excepcionalmente se llegara a evaluar en forma censal a todos los alumnos del país, o de una de sus entidades, en ciertas materias y ciertos grados clave, nunca sería adecuado que las decisiones de alto impacto sobre una persona, como las que se refieren a su aprobación o reprobación, o a la decisión de admitirlo o no, en un grado o ciclo superior, se tomara únicamente con base en los resultados que obtuviera en pruebas en gran escala.

Las evaluaciones individuales de alto impacto siempre deben basarse en la evidencia más amplia de que pueda disponerse, incluyendo la ofrecida por los



juicios evaluativos de los maestros y la escuela involucrada, cuyo peso en la decisión final nunca debería ser menor en relación con los resultados eventuales de pruebas en gran escala. Debe añadirse que el contar con buenas evaluaciones sobre muestras nacionales de un Sistema Educativo permite también valorar la calidad de los otros mecanismos de evaluación, siempre con el propósito de mejorar la educación en conjunto.

El subsistema de evaluación de escuelas

Este tercer componente del INEE tendrá como objetivo apoyar a las autoridades correspondientes en la evaluación de las escuelas como organizaciones singulares, para lo cual diseñará modelos de autoevaluación de planteles y sistemas de evaluación y monitoreo externo complementarios. Se podrá incluir el desarrollo de instrumentos para apoyar la evaluación de habilidades, conocimientos, actitudes, valores y desempeño del personal docente, el de apoyo y el de quienes realicen funciones de dirección y supervisión en el sistema de educación básica.

Las dimensiones del concepto de calidad que se utilizan están construidas poniendo en relación las necesidades del entorno con los objetivos y los productos del sistema, y éstos con los insumos del mismo.

El componente central, constituido por los procesos, no forma parte, como tal, de las definiciones de relevancia, eficacia, impacto, equidad y eficiencia. Sin embargo, la importancia de tener en cuenta los procesos en una evaluación de la calidad de un sistema educativo es clara. Los procesos darán lugar a los productos, a partir de los insumos. Son los procesos los que funcionan bien o mal y se deben perfeccionar o corregir para que los productos sean mejores.

Por otra parte, los procesos no se dejan capturar mediante indicadores numéricos con tanta facilidad como los insumos y los productos. Por el contrario, usualmente el tratamiento de los procesos debe ser eminentemente cualitativo, como se aprecia en los sistemas desarrollados para atender estos aspectos, como los tradicionales inspectorados, al igual que mecanismos recientes como el GRIDS inglés, el QUAPE catalán o el programa americano de escuelas excelentes, llamadas de *etiqueta azul* (*blue ribbon schools*). Para ser completo, un sistema de evaluación

deberá comprender un componente de este tipo, del que se derivará la información necesaria para generar indicadores de proceso.

Reiterando una idea expresada en el apartado anterior, conviene hacer una consideración más, a propósito del subsistema de evaluación de escuelas: una de las orientaciones generales que se proponen para el INEE establece que los resultados de sus evaluaciones no deberán referirse a individuos, sino a sistemas o subsistemas.

Como se mencionó, esto se aplica, sin duda, a los alumnos y a los profesores: el trabajo basado en muestras representativas, las características técnicas de los instrumentos que siempre hacen necesarios otros elementos para la toma de decisiones de alto impacto, junto a consideraciones de equidad, dejan claro que las decisiones sobre alumnos individuales (aprobación o reprobación, promoción y similares) no deban hacerse a partir de las evaluaciones del INEE, sino con base en otros mecanismos de evaluación que se manejan en el nivel de cada escuela; lo mismo puede decirse en cuanto a las decisiones relativas a los maestros en lo individual, como las concernientes con su contratación, basificación, salario, estímulos y, eventualmente, sanciones.

Reiterando la idea esencial: las evaluaciones del INEE nunca deberán usarse para tales propósitos, si bien podrán servir para calibrar los mecanismos de evaluación que se utilicen para ello en cada escuela, zona escolar o entidad.

En lo que se refiere a las escuelas individuales, en cambio, una escuela es, en cierto sentido, un microsistema educativo. Por ello no se descarta *a priori* que el trabajo evaluativo del INEE pueda llegar a servir de base para decisiones sobre escuelas en lo individual, si bien esto deberá hacerse siempre con cuidado, nunca como elemento único, sino en forma conjunta con otros elementos, con base en criterios y procedimientos previamente establecidos con claridad y, según las orientaciones generales, siempre con propósitos de mejoramiento.

Los tres subsistemas del INEE se integrarán en un todo perfectamente articulado que, con las características jurídicas y organizacionales adecuadas y en estrecha relación con la autoridad educativa federal (SEP) y las de las entidades (las secretarías de educación de los estados o sus equivalentes), así como con



- Las tres columnas corresponden a otros tantos *elementos* del Sistema Educativo, a saber sus insumos, sus productos y los procesos que tienen lugar en su interior.
 - Los cuatro renglones (*micro, meso, macro y mega*), corresponden, a su vez, a otros tantos *niveles* de dicho sistema, desde el más pequeño, o de aula, hasta el mayor, del sistema mismo, considerado como un todo.
- En cada uno de los renglones se consideran, de manera simplificada, dos tipos o niveles del Sistema Educativo, el de la educación básica obligatoria (que se designa con las iniciales EBO), y el de la educación post-obligatoria (EPO), el cual corresponde a la llamada media superior y la superior. En cada una de las casillas del cuadro, definida por el cruce de columnas y renglones, se precisan tipos de evaluación específica.

CUADRO I. COMPONENTES DEL SISTEMA NACIONAL DE EVALUACIÓN EDUCATIVA

Niveles de evaluación, tipos y niveles educativos, instancias	ELEMENTOS DEL SISTEMA EDUCATIVO		
	INSUMOS Indicadores	PRODUCTOS (Aprendizaje)	PROCESOS (Pedagógicos y organizacionales)
MICRO. EBO: director y maestros de una escuela de educación básica EPO: autoridades-académicos de inst. Chica, dependencia o programa	Generan información Usan resultados Generan información Usan resultados	Evaluación primaria del aprendizaje de alumnos Evaluación primaria del aprendizaje de alumnos	Hacen autoevaluación de la escuela Autoevaluación de la institución, de la dependencia o del programa
MESO. EBO: supervisores de zona escolar EPO: Instituciones de Ed. media sup. O Superior	Integran información Usan resultados Generan información Usan resultados	Supervisión de prácticas de evaluación primaria Certificación alumnos	Evaluación externa de primer nivel; supervisión Autoevaluaciones institucionales
MACRO. EBO: Áreas estatales evaluación EPO: COEPES, COPEMS	Integran información, usan resultados, hacen indicadores propios Integran información, usan resultados, hacen indicadores propios	Certificación alumnos Análisis eval. nacionales Evaluaciones estatales Supervisión de IES Análisis eval. nacionales Evaluaciones estatales	Eval. externa de segundo nivel Evaluación externa de primer nivel
MEGA. EBO: SEP (DGE, DGPPP), INEE EPO: Ceneval, CIEES, Copaes, Organismos acreditadores	Diseña/mantiene sist. nacional indicadores Análisis nacionales Diseña/mantiene sist. nacional indicadores Análisis nacionales	Diseño de instrumentos Evaluaciones nacionales Diseño de instrumentos Evaluaciones nacionales	Diseñan metodologías, capacitan a usuarios otros niveles Eval. externa de segundo nivel; diseño metodologías, capacitación

El *Programa Nacional de Educación 2001-2006*, en el apartado 3 de la primera parte, en la que se resume la situación del Sistema Educativo Nacional, destaca la importancia de la evaluación y reconoce los avances previos, pero considera que aún no puede hablarse de un verdadero sistema nacional de evaluación:

La evaluación permanente y sistemática, que combina la participación de instancias externas e internas, es un importante instrumento de gestión. Permite contar con información que es útil para llevar a cabo acciones de mejora, así como para rendir cuentas a la sociedad. En la década pasada, la noción de evaluación comenzó a ser aceptada; la cultura de la evaluación, con todo, aún es frágil entre los responsables del sistema educativo. Éste ha sido uno de los motivos de que pocas evaluaciones se hayan hecho públicas.

Hoy existen instituciones y mecanismos dedicados a la evaluación en todos los niveles educativos, y ha comenzado la colaboración entre los responsables de las funciones asociadas con la evaluación en las 32 entidades federativas. Pese a ello, todavía no se puede hablar de un auténtico sistema nacional de evaluación educativa, capaz de hacer frente a la tarea que le corresponde, ni hay suficientes especialistas para asumirla. (2001: 69)

A partir del diagnóstico anterior, en el apartado que contiene la visión del Sistema Educativo Nacional a 2025, que designa con la expresión *un enfoque educativo para el siglo XXI*, y en lo que se refiere a la calidad y su evaluación, el PNE establece:

Los niveles de aprendizaje de los alumnos se habrán elevado, al tiempo que las diferencias regionales se habrán reducido y la calidad del Sistema Educativo será reconocida nacional e internacionalmente. Lo anterior podrá apreciarse gracias a mecanismos rigurosos y confiables de evaluación de todos los tipos y niveles, cuyos resultados se difundirán y utilizarán para el mejoramiento y serán altamente valorados por maestros, directivos y la sociedad en general. Todas las instituciones públicas y privadas satisfarán criterios básicos de calidad, acordes a su naturaleza. Con apoyo en un eficaz sistema de información, la evaluación y la investigación educativa apoyarán la toma de decisiones y el rendimiento de cuentas. (2001: 73)

Para avanzar en la dirección a la que apunta la visión a 2025, el PNE establece objetivos precisos en el horizonte de la presente administración federal, a 2006. Los relativos a evaluación se agrupan en el objetivo particular 5 de la segunda parte:

Fortalecer el funcionamiento del Sistema Educativo Nacional, mediante la consolidación del sistema de evaluación, el fomento de la investigación y la innovación educativa, renovados sistemas de información e indicadores. (p. 100)

El objetivo particular anterior se desglosa en cinco conjuntos de líneas de acción, el primero de las cuales se designa precisamente con la expresión *consolidar el Sistema Nacional de Evaluación Educativa*, e incluye siete líneas de acción particulares:

- Crear el Instituto Nacional para la Evaluación de la Educación, cuyas funciones se articularán con las de las instancias responsables de la evaluación de la SEP y las entidades federativas.
- Crear un organismo para la evaluación de las instituciones de educación media superior.
- Promover la evaluación de todos los programas educativos que se ofrecen en el Sistema de Educación Superior.
- Fomentar la creación de organismos independientes para la acreditación de programas e instituciones en áreas diversas del conocimiento y del ejercicio profesional.
- Promover la consolidación del Ceneval, los CIEES y el Copaes.
- Promover la articulación de los diversos mecanismos de evaluación de la educación superior.
- Establecer lineamientos para asegurar la calidad de las evaluaciones y su adecuada utilización y difusión, evitando que se produzcan efectos contrarios a la equidad.

El tercer conjunto de líneas de acción del objetivo 5 se denomina *Fortalecer la cultura de la planeación y la evaluación*, e incluye dos que establecen:

- Se diseñarán e instalarán sistemas innovadores de planeación y evaluación de los procesos, proyectos, programas y políticas que ejecuta la SEP a través de las áreas centrales y los organismos paraestatales, y se propondrá su aplicación en los proyectos que ejecutan las entidades federativas.
- Se transitará de una concepción de la evaluación como mecanismo de control o fiscalización a la evaluación como un medio fundamental para propiciar aprendizajes individuales y organizacionales.

Entre las líneas de acción del cuarto conjunto, *De-*



sarrollar el Sistema Nacional de Indicadores Educativos, se encuentran una que establece expresamente:

- Concebir el nuevo Sistema de Indicadores Educativos como el elemento que articulará e integrará los resultados del aprendizaje de los alumnos con los de la gestión de escuelas, subsistemas y el sistema educativo mismo, de manera que haya disponibilidad permanente de información relevante para la toma de decisiones en los diferentes niveles de gestión de la educación.

Puede apreciarse que el *Programa Nacional de Educación* tiene objetivos, líneas de acción y metas precisas, que se refieren precisamente a elementos de un Sistema Nacional de Evaluación Educativa de concepción más amplia al que comenzó a desarrollarse en el período 1995-2000.

Para establecer un *Sistema Nacional de Evaluación Educativa* muy sólido idealmente serían convenientes algunos ajustes al marco jurídico de la educación nacional, y en particular a la Ley General de Educación, a la legislación relativa a educación superior y, posiblemente, a la Constitución, para dar cabida a los elementos del SNEE, de manera más acorde con las concepciones actuales sobre evaluación educativa, y con las que se refieren al fortalecimiento de la democracia, la llamada *ciudadanización* de la vida pública, la transparencia y la rendición de cuentas.

Asimismo, convendría actualizar la Ley Orgánica de la Administración Pública Federal y la Ley Federal de Entidades Paraestatales, y de esa manera puedan crearse organismos públicos dotados de independencia técnica, que el marco jurídico vigente no contempla, sin llegar al rango de constitucionalmente autónomos.

Sin embargo, los procesos legislativos, sobre todo si implican cambios al máximo ordenamiento, son necesariamente a largo plazo.

Por ello parece conveniente dar pasos en la dirección de un vigoroso Sistema Nacional de Evaluación Educativa procediendo por una vía más rápida, en tanto se lleva a cabo el proceso más largo de consultas y consensos será necesario para la adecuación del marco jurídico nacional, en la dirección apuntada. Los pasos precisos que parecería conveniente dar en el corto plazo son:

- Crear, por la vía que se considere más expedita, el Instituto Nacional para la Evaluación de la Edu-

cación (INEE), como organismo público descentralizado que apoye a la SEP, a las autoridades educativas estatales y a la sociedad en general, en lo relativo a evaluación diagnóstica, para sustentar decisiones estratégicas de política educativa, tanto para el Sistema Educativo Nacional, como para los subsistemas de las entidades federativas, en cuanto tales.

- Crear el organismo que contempla el *Programa Nacional de Educación 2001-2006* para la acreditación de los planteles de educación media superior.
- Por lo que se refiere a la educación superior, avanzar en la consolidación del Ceneval y los Comités de Pares (CIEES), contemplando su posible integración en un organismo similar al INEE, al que pueda añadirse un subsistema fortalecido de indicadores de calidad de la educación superior, y considerando también la consolidación del Copaes.

Si la Ley General de Educación se enriqueciera con un capítulo en el que se defina la estructura y funciones del Sistema Nacional de Evaluación en lo relativo a educación básica, y se legislase de manera similar en lo relativo a educación superior, podrá definirse una relación del SNEE y sus componentes con otras instancias, como el Consejo Nacional de Participación Social en Educación (Conapase: al que la Ley asigna la función de recibir los resultados de las evaluaciones), los Consejos de Vinculación para los niveles medio superior y superior (Cvems, Cves), el Consejo Nacional de Autoridades Educativas (CNAE) y el Consejo de Especialistas en Educación (CEE).

Si la legislación precisara las funciones de esos cuerpos colegiados en cuanto a la definición de los grandes lineamientos de política educativa, ello debería comprender lo relativo a evaluación, como un elemento importante de la estrategia global de calidad. De esta manera podría lograrse una articulación precisa de las acciones de evaluación, desarrolladas por los diversos elementos del SNEE, con el conjunto de las políticas educativas, sin que esto implicara una dependencia jerárquica directa con respecto a las autoridades educativas federales. Con un buen nivel técnico, lo anterior sería una sólida base para la credibilidad pública de la evaluación educativa.

La estructura del SNEE, incluyendo el nivel nacional y el estatal de los tres grandes tipos de educación, se presentaría de la siguiente manera:

CUADRO 2. ESTRUCTURA DEL SISTEMA NACIONAL DE EVALUACIÓN EDUCATIVA

NIVELES	Conapase, CVEMS, CVES, CNAE, CEE		
	Educación Básica	Ed. Media Superior	Educación Superior
Sistema Educativo Nacional	INEE	Nuevo organismo	Ceneval, Copaes, CIEES, O. acreditadores
Subsistemas Estatales	Áreas Estatales de Evaluación	COPEMS	COEPES

3.6 La evaluación como parte de una estrategia global de calidad

Lo más importante con respecto a la calidad educativa no es conceptualizarla ni medirla, sino mejorarla. Conceptualización y medición tienen sentido en la medida en que aporten elementos para el mejoramiento, sin ellos se dificulta. Para su mejoramiento hace falta, además, como se ha indicado, saber a qué se debe que la calidad esté presente o ausente, aumente o disminuya; en otras palabras, se necesita contar con explicaciones y comprensiones adecuadas de los elementos que inciden en la calidad educativa, eso permitirá definir estrategias para mejorarla cuando se crea necesario.

Para propósitos de mejoramiento la descripción de un sistema educativo no basta; hacen falta explicaciones que se acerquen a las causas de la situación descrita. La detección de los factores del rendimiento escolar, la valoración del peso de cada uno, el conocimiento de su *modus operandi* y de la forma en cómo interactúa con los demás, de eso se habló en el apartado 3.2 de este documento; permitirán que los tomadores de decisiones cuenten con bases sólidas para diseñar estrategias adecuadas de mejoramiento.

Si se tienen en cuenta los muchos factores que inciden en el aprendizaje de los alumnos y, por ende, en la calidad educativa, se aceptará el planteamiento siguiente, en el sentido de que, para mejorar la calidad educativa sería necesario idealmente un amplísimo conjunto de esfuerzos donde se incidiera en todos esos factores. Para designar ese conjunto de acciones de mejoramiento utilizaremos la expresión *estrategia global de calidad*, que incluye varias estrategias particulares, esta se sistematiza de la manera siguiente:

- *Estrategias curriculares:* este primer componente de la estrategia global tiene que ver con la primera dimensión de la noción de calidad, la *Relevancia*, y también con algunos insumos. Incluye fundamentalmente lo relativo a la formulación de un currículo cuyos objetivos sean congruentes con las necesidades de la sociedad en que se ubica el sistema educativo, pero también la elaboración de libros de texto y otros materiales de apoyo. Supone un trabajo de diseño curricular que, por una parte, analice cuidadosamente las necesidades del entorno, en todas sus dimensiones (las económicas pero también las sociales, culturales y políticas) y, por otra, que aproveche los mejores avances de las ciencias cognitivas y la tecnología educativa; incluye el cuidadoso establecimiento de estándares de aprendizaje.
- *Estrategias relacionadas con los maestros:* a partir de la convicción del papel central que juega el magisterio en los procesos educativos, una estrategia global deberá dar gran importancia a la elevación de la calidad de este componente del sistema. Se deberá incluir lo relativo a la formación inicial de los docentes (y por lo tanto el mejoramiento de las escuelas normales), pero también la actualización permanente, el fomento de la profesionalización y el aseguramiento de condiciones de trabajo adecuadas incluyendo, por supuesto, lo relativo a salarios y estímulos.
- *Estrategias relacionadas con las escuelas:* considerando que es en ellas el único lugar en donde puede darse la mejora de la calidad, en este rubro deberá contemplarse todo lo concerniente con las condiciones de infraestructura y equipamiento pero, sobretudo, con las condiciones organizacionales, en especial el liderazgo del director, la



actuación del equipo docente, la existencia de un proyecto escolar propio, que se proponga objetivos ambiciosos, a partir de un diagnóstico realista de la situación local, todo lo cual presupone un grado razonable de autonomía.

- *Estrategias de apoyo por parte de las estructuras intermedias:* desde la zona escolar y los inspectores hasta el gobierno federal, pasando por el nivel estatal. De importancia especial parecen la revitalización de la supervisión (cuya función teórica de apoyo pedagógico y evaluación, se ha desvirtuado en aras de lo administrativo y sindical) y los programas compensatorios. La experiencia ha mostrado la ingenuidad de las expectativas suscitadas por los primeros programas de este tipo, pero también que, sin ser soluciones milagrosas, pueden representar una diferencia significativa para los niños pobres si se les destinan recursos, si su diseño tiene en cuenta los problemas específicos que dificultan el aprendizaje y si se cuida la implementación.
- *Estrategias de apoyo por parte de la sociedad:* a partir de la consideración de que la educación y su calidad son asunto de todos, y no sólo de las autoridades educativas, este rubro incluirá centralmente lo relativo a la participación de los padres de familia en el funcionamiento de la escuela de sus hijos y en el apoyo al aprendizaje de estos, pero también las acciones para involucrar a los demás sectores de la sociedad en el apoyo a la tarea de la escuela. Pueden mencionarse en especial a las empresas y a los medios masivos de comunicación, pero ningún sector de la sociedad debería quedar al margen de este esfuerzo colectivo.
- *Estrategias de evaluación:* en el marco de la estrategia global los esfuerzos de evaluación de la calidad tienen un lugar importante, cuyas peculiaridades se desarrollan a continuación.

Además de los anteriores componentes de la estrategia global de calidad, ésta deberá tener en cuenta, por lo menos otros dos elementos, que no son parte de ella en sentido estricto, pero sí prerequisites indispensables: por una parte, la existencia de un marco jurídico adecuado (por ejemplo para que cada escuela tenga el mínimo de autonomía necesario para su funcionamiento como unidad organizacional); el otro elemento de

este tipo es la existencia del financiamiento mínimo que permita el funcionamiento adecuado de las escuelas y el sistema educativo en su conjunto.

Todo lo referido soporta la afirmación de que la evaluación, como la calidad del sistema educativo en general, debe concebirse como una gran tarea colectiva, de la que ningún actor importante debe quedar marginado. Reiterando, de alguna manera, lo dicho hasta ahora, los actores cuyo concurso no debe faltar en un buen sistema de evaluación son, por lo menos:

- Los alumnos que, aunque muchas veces parecen concebirse como objeto de las políticas y acciones, son sin duda los sujetos educativos fundamentales, a cuyo desarrollo deben orientarse los esfuerzos de todos los demás actores.
- Los maestros, cuyo trabajo profesional cotidiano es la acción esencial, imprescindible para que exista la escuela y pueda darse en ella la calidad.
- Los directores que, en las escuelas con varios grupos de alumnos, atendidos por otros tantos maestros, deben aportar el liderazgo y la conducción del trabajo de equipo esto propiciará que hará que la escuela se asuma como entidad con una dinámica propia, y la haga merecedora de ser calificada como escuela de buena calidad.
- Los padres de familia, además del apoyo personal que puedan ofrecer a sus hijos, deben ser miembros activos de la comunidad escolar, con derechos y obligaciones que defina una legislación más adecuada a la actual, y apoyen al equipo docente, con el liderazgo activo y prudente de un buen director.
- Los académicos e intelectuales y, en particular, los investigadores que estudian la educación, quienes pueden aportar elementos valiosos, tanto para el diseño de políticas educativas efectivas, como para una evaluación rigurosa de los resultados educativos.
- Todos los sectores de la sociedad, en especial: los empresarios, sobre todo en relación con la educación media superior y la superior; los medios masivos de comunicación, por el peso que tiene su influencia en la sociedad contemporánea y en la población en edad escolar; y las ONG que tienen a la educación como objeto de interés.
- Las autoridades educativas municipales, estatales y federales, cuya acción puede ser decisiva

- El poder legislativo, en cuyas manos está la facultad de establecer un marco jurídico adecuado

y asignar recursos suficientes en los presupuestos públicos, también los requisitos señalados como necesarios para que el sistema educativo pueda funcionar y alcanzar los niveles de calidad deseados.

4.1. Marco jurídico: aspectos técnicos y toma de decisiones en evaluación.

Art. 73, XXV: ...dictar las leyes encaminadas a distribuir convenientemente entre la Federación, los Estados y los Municipios el ejercicio de la función educativa y las aportaciones económicas correspondientes a ese servicio público, buscando unificar y coordinar la educación en toda la República...

La Ley General de Educación precisa tal distribución de facultades entre los niveles de gobierno. En lo relativo a evaluación, esta Ley incluye las siguientes disposiciones:

CAPÍTULO II. DEL FEDERALISMO EDUCATIVO

I. Determinar para toda la República los planes y programas de estudio para la educación primaria, la secundaria, la normal y demás para la formación de maestros de educación básica...

XI. Realizar la planeación y la programación globales del sistema educativo nacional, evaluar a éste y fijar los lineamientos generales de la evaluación que las autoridades educativas locales deban realizar;

Sección 4. De la evaluación del Sistema Educativo Nacional.

Artículo 30. *Las instituciones educativas establecidas por el Estado, por sus organismos descentralizados y por los particulares con autorización o reconocimiento de validez oficial de estudios, otorgarán a las autoridades educativas todas las facilidades y colaboración para la evaluación a que esta sección se refiere. Para ello, proporcionarán oportunamente toda la información que se les requiera; tomarán las medidas que permitan la colaboración efectiva de alumnos, maestros, directivos y demás participantes en los procesos educativos; facilitarán que las autoridades educativas, incluida la Secretaría, realicen exámenes para fines diagnósticos y recaben directamente en las escuelas la información necesaria.*

Artículo 31. *Las autoridades educativas darán a conocer a los maestros, alumnos, padres de familia y a la sociedad en general los resultados de las evaluaciones que realicen, así como la demás información global que permita medir el desarrollo y los avances de la educación en cada entidad federativa.*

Para aplicar adecuadamente las disposiciones anteriores son necesarias algunas precisiones conceptuales. Debe subrayarse, en especial, que la evaluación no puede limitarse a una descripción, por detallada



que sea, de cierta realidad; en el caso de la evaluación educativa, la del nivel de aprendizaje alcanzado por los alumnos de una institución o sistema escolar.

La *evaluación* requiere, ciertamente, de descripciones y mediciones precisas, pero no puede reducirse a ellas; evaluar implica como elemento constitutivo un juicio valorativo mediante el cual se compara una situación existente –por ejemplo el nivel de aprendizaje alcanzado– con ciertos estándares de calidad o parámetros de referencia.

Los *estándares* no se definen a partir de datos empíricos, por precisos que sean, sino que se establecen con base en elementos normativos –en educación los objetivos del currículo– que, a su vez, se derivan de elementos normativos, jurídicos y filosóficos más amplios, como las ideas sobre lo que debe saber una persona para cumplir adecuadamente con sus responsabilidades como ciudadano o, en su caso, como profesionista o persona especializada en cierto ámbito.

La responsabilidad de evaluar el sistema educativo entendido de esta manera, no puede recaer en una instancia diferente de la misma autoridad educativa, que tiene la facultad de definir el currículo y de manejar al sistema mismo. El indispensable balance entre autoridad y responsabilidad que debe darse en toda institución así lo requiere.

Pero nada exige que las actividades técnicas de recolección de información empírica sobre el aprendizaje alcanzado por los alumnos y otros aspectos del sistema educativo susceptibles de evaluación, deban ser realizadas directamente por la misma autoridad educativa. Es perfectamente concebible que las tareas técnicas relativas a la evaluación, que designaremos genéricamente como medición, sean realizadas por instancias especializadas independientes, que entreguen sus resultados a la autoridad. Esas tareas incluyen:

- La de diseñar los instrumentos necesarios para recabar la información empírica, con características que aseguren su validez y confiabilidad.
- La de aplicarlos en forma que asegure la calidad de la información obtenida.
- La de sistematizar los datos así recopilados y analizarlos de manera que puedan servir de base para la toma de decisiones.
- La de hacer comparaciones transversales y longitudinales, nacionales o internacionales.

- Y la de realizar investigaciones específicamente orientadas para entender las causas de las tendencias y diferencias que se encuentren.

Como muestra la experiencia internacional, la complejidad técnica de las tareas de medición y las relacionadas, así como la necesidad de continuidad y credibilidad, hacen conveniente que estén a cargo de órganos independientes.

Estos órganos pueden realizar todas las tareas técnicas mencionadas, incluyendo la entrega de resultados a las autoridades y su difusión en beneficio de otras instancias legítimamente interesadas, como las asociaciones de los padres de familia y la sociedad en general.

La evaluación misma, en cambio –el juicio valorativo que, comparando la información empírica con los parámetros normativos, establece lo que está bien y mal, lo que debe estimularse qué amerita apoyo o corrección– no es forzosamente, en sentido estricto, facultad de la misma instancia que se encarga de la medición.

Al hacer comparaciones longitudinales, o transversales, el organismo responsable de la medición evalúa, de alguna manera, como también lo hace cualquier ciudadano, quien puede contrastar los resultados que se difunden de las mediciones con sus propias expectativas. Pero la evaluación en el sentido más estricto, el único que tiene valor legal, quien compara los resultados de las mediciones con los estándares oficiales y decide, por ejemplo, la promoción de un alumno al ciclo o grado siguiente o su reprobación, no puede corresponder sino a la autoridad educativa.

La evaluación propiamente dicha, corresponde invariablemente a la misma instancia que tiene la facultad de establecer, en el currículo, los parámetros de referencia, los objetivos del sistema educativo, cuyo logro será posible valorar gracias a la información empírica obtenida.

Sin embargo, como se mencionó, en la medida en que el INEE difunda directamente los resultados obtenidos por los diversos subsistemas educativos nacionales, permitiendo compararlos entre sí, en un momento dado y a lo largo del tiempo, y en la medida en que su trabajo permita comparar los resultados educativos mexicanos con los de los sistemas de otros países, gracias al uso de instrumentos



de medición equiparables, habrá un componente estrictamente evaluativo en el quehacer del Instituto, el cual no podrá sino ser percibido como tal por el público.

Por otra parte, conviene reiterar que el tipo de evaluación al cual se refiere la Sección 4 del Capítulo II de la Ley General de Educación, es la evaluación del Sistema Educativo Nacional o, en su caso, de los sistemas estatales, y no la de los alumnos y los maestros en lo individual.

Este punto debe ser comprendido cabalmente para que pueda precisarse la función del INEE, en relación con la de la SEP, en el seno del Sistema Nacional de Evaluación Educativa. La evaluación del sistema nacional y los subsistemas, estatales u otros, implicará, como se ha dicho, la obtención de datos sobre los niveles promedio de aprendizaje de los alumnos y, eventualmente, sobre otros aspectos de alumnos y maestros, datos que se podrán obtener de todo el universo o, por lo general, de muestras representativas.

Como se mencionó, la información así obtenida no deberá utilizarse para hacer evaluación de individuos, sino únicamente para la evaluación del sistema nacional, de subsistemas estatales u otros, como las escuelas del medio urbano en contraparte con la rural, etcétera. Con las debidas precauciones, y los elementos del tercer componente del sistema, podrá llegarse hasta la evaluación de escuelas.

En síntesis: la SEP —y, en la medida en que la Ley lo indica, las autoridades educativas locales— son las instancias responsables de la conducción del Sistema Educativo Nacional y los subsistemas estatales, y sólo a ellas corresponde la determinación de los objetivos curriculares de la educación básica y de los parámetros que deberán utilizarse en un sistema de indicadores de calidad.

Sólo puede corresponder a esas autoridades, en consecuencia, la evaluación, en sentido estricto y completo, de la consecución de los objetivos y, por supuesto, las decisiones subsecuentes de aprobación o reprobación, estímulo o sanción. Pero nada impide que se encomiende a una instancia relativamente autónoma, como se quiere sea el INEE, la realización de las tareas técnicas necesarias para la medición de los aprendizajes y la obtención de información válida y confiable sobre otras variables relevantes para la

evaluación educativa; con base en la experiencia internacional se afirma que ello es, recomendable. Las funciones del INEE, en cuanto organismo independiente, podrán incluir otras acciones relacionadas con la evaluación, como las que implican las comparaciones transversales y longitudinales la cual la publicación de resultados permite, tanto entre subsistemas estatales como en el plano internacional.

Para precisar la relación entre la SEP y las autoridades educativas de las entidades, por una parte, y el INEE por otra, y para que las funciones de las partes se complementen y articulen armoniosamente, en un verdadero Sistema Nacional de Evaluación Educativa, será necesario precisar las normas generales de la Ley General de Educación, ya citadas, en una Política Nacional de Evaluación del Sistema Educativo (PNESE), con la que deberán ser congruentes las funciones del INEE, y los demás componentes del SEN.

La PNESE deberá tener valor legal, y se precisará y ajustará a partir de las necesidades que presente la evolución del sistema educativo; de igual forma deberán precisarse las funciones del INEE, y los términos de referencia de sus trabajos, en relación con las actividades y metas del Sistema Educativo Nacional.

La *Política Nacional de Evaluación del Sistema Educativo* deberá contener lineamientos precisos relativos, por lo menos, a los puntos siguientes:

- Los puntos de referencia con los que se deberán comparar los resultados obtenidos para llegar a juicios de valor sobre la calidad educativa, tanto en una perspectiva transversal como en una longitudinal, por ejemplo países de la OCDE, países de América Latina, situación anterior del sistema, situación futura deseable en términos de metas, etcétera.
- Las consecuencias de la evaluación, en términos de apoyos compensatorios, estímulos, medidas preventivas y correctivas, y financiamiento.
- La difusión pública de los resultados de la evaluación, cuidando tanto el justo derecho de las personas a la privacidad, como el de la sociedad a que se le rindan cuentas sobre el uso de los recursos públicos y el funcionamiento de servicios de interés general.
- La distinción entre la evaluación de personas, la de instituciones y subsistemas, y la del Sistema Educativo Nacional en su conjunto.

Éstos y otros lineamientos de la PNESE deberán buscar que la actividad del INEE tenga consecuencias positivas para la calidad educativa, evitando comparaciones injustas de las que puedan derivarse inequidades en los apoyos financieros, efectos indeseables en la opinión pública o en los maestros, como la posible tendencia de éstos a concentrarse en la preparación inmediata de sus alumnos para las pruebas.

4.2. Rasgos básicos del INEE

Además de su estructura básica tripartita, los rasgos fundamentales del INEE pueden resumirse diciendo que el organismo deberá tener:

Una base conceptual rica: La estructura y la operación del INEE deberá partir de un concepto multidimensional de calidad, que incluya las dimensiones de relevancia, eficacia interna y externa, impacto, eficiencia y equidad.

Una cobertura completa: En una perspectiva de mediano y largo plazo, los tres subsistemas del INEE deberán cubrir todos los niveles y grados de la educación básica; todas las entidades de la República y las modalidades y variantes de la oferta de servicios; todas las áreas del currículo y los niveles de operaciones cognitivas.

Una función técnica: Reconociendo la responsabilidad de las autoridades para la evaluación de los sistemas educativos nacional y estatales, el INEE desempeñará una función técnica de medición, como apoyo a dichas autoridades y a la sociedad.

Una función evaluativa: en sentido estricto, en la medida en que así lo estipule la *Política Nacional de Evaluación*, especialmente en lo relativo a comparaciones transversales y longitudinales nacionales e internacionales.

Un papel estratégico: derivado de las funciones anteriores, ya que la información que genere deberá ser aprovechada por las autoridades para mejorar la educación.

Una función de difusión y apoyo a la rendición de cuentas: Como la educación es de interés público, el trabajo del INEE deberá incluir una actividad de comunicación, en dirección de

los poderes legislativo y judicial, así como de los maestros, los padres de familia, los medios masivos de comunicación y la sociedad en general, para fomentar el desarrollo de una cultura de la evaluación.

Una función de investigación: Para hacer su tarea con el mejor nivel, el INEE deberá tener una función de investigación, tanto para el avance de la metodología de evaluación, como para explicar sus hallazgos. El INEE realizará investigaciones por sí mismo y promoverá su realización por otras instancias, celebrando convenios y abriendo sus bases de datos a investigadores independientes, con normas claras.

Una función de capacitación: Para que en todas las entidades haya personal calificado para realizar actividades de evaluación propias y para interpretar y utilizar las evaluaciones nacionales, en apoyo a las políticas de mejoramiento, y ante la escasez de especialistas en el campo, en colaboración con la SEP el INEE deberá desarrollar actividades de formación y capacitación diversas: enviando becarios a buenos posgrados, organizando programas acelerados de formación básica, e impulsando programas académicos en colaboración con universidades.

Una base institucional apropiada: Para que los puntos anteriores puedan ser realidad, el INEE deberá tener una estructura adecuada, la cual asegure su autonomía en relación con el ejecutivo y cualquier instancia externa, a partir de su definición como institución técnica y académica. Para ello deberá tener órganos colegiados de dirección con participación equilibrada de autoridades, especialistas y representantes de actores importantes del sistema educativo, y contar con directivos y personal profesional de base, seleccionado con base en criterios académicos.

4.3. Orientaciones generales

Con mucha frecuencia, la ausencia de una cultura de evaluación en los medios educativos mexicanos trae como resultado, una inadecuada comprensión de los

En el desarrollo de sus funciones, el Instituto buscará siempre contribuir al mejoramiento integral de la educación mexicana, mediante una contribución específica, en el marco de las grandes orientaciones que señalan la Constitución de los Estados Unidos Mexicanos y la Ley General de Educación. En particular, el Instituto se regirá por las siguientes orientaciones generales:

- i) *Procurará que los resultados de sus evaluaciones permitan comparar en el tiempo.*
- j) *Se esforzará por participar en todos los esfuerzos pertinentes de evaluación internacional (véase coordinación del Área Educativa, 2000: 160-161).*

En lo referente al alto nivel técnico que deberá mantener en todas sus actividades, asimismo el INEE atender los criterios de calidad internacionalmente aceptados por los organismos especializados del campo.

Personalidad jurídica

Según la Constitución y la Ley Orgánica de la Administración Pública Federal (Art. 45), hay dos vías para crear un organismo público descentralizado: por decreto administrativo del titular del ejecutivo, o mediante decreto o Ley del Congreso. En el segundo caso puede tratarse de organismos sectorizados, cuya dirección depende del titular de una secretaría, de entidades desectorizadas e, incluso de organismos constitucionalmente autónomos como el Instituto Federal Electoral, el Banco de México y la Comisión Nacional de Derechos Humanos. Hay, además, la posibilidad de constituir una Asociación Civil.

Después de considerar las ventajas y desventajas de estas opciones, así como los pasos a dar para implementar cada una, se llega a la conclusión de que, aunque la evaluación educativa es una función importante, el *estatus* de organismo constitucionalmente autónomo parece excesivo para una instancia técnica y académica como la que se propone; en el otro extremo, el carácter de Asociación Civil resulta inadecuado para un organismo que desarrollará una función pública tan cuidadosa. Por ello las opciones que se creen adecuadas se reducen a las de órgano público



descentralizado, sea creado por ley del Congreso, sea por decreto presidencial.

En ambos casos el INEE quedaría sujeto a las disposiciones administrativas aplicables a todo organismo público, incluyendo las relativas a los procedimientos de contraloría interna y externa tanto administrativa como financiera. El tener personalidad propia le permitirá quedar fuera de la estructura de la SEP, permitiendo una organización propia y facilitando un funcionamiento eficiente.

Para crear un organismo constitucionalmente autónomo se requiere obviamente de una modificación de la máxima ley; si se le crea como organismo no sectorizado el Secretario de Educación no presidirá el órgano de gobierno, y en éste podrá lograrse un equilibrio mejor de funcionarios federales y sectores sociales y académicos; si se le establece como organismo sectorizado, por ley o por decreto presidencial, el titular del ramo deberá presidir el órgano de gobierno, cuya composición deberá ser más limitada.

En cualquier concepto será necesario buscar la mayor credibilidad para el nuevo organismo, lo cual implica que cuente con un Consejo Técnico fuerte, al cual competan las decisiones de tipo técnico-académico; en tanto el órgano de gobierno se limite a las administrativas.

Por otra parte, es posible pensar en que el Instituto sea creado por decreto presidencial, por ser la forma más rápida de hacerlo, y posteriormente el Congreso modifique su personalidad jurídica para que goce de mayor autonomía, sea de rango constitucional, sea sólo académica y técnica, lo cual se facilitaría con modificaciones de la Ley General de Educación y la Ley Federal de Entidades Paraestatales.

Estructura operativa

La estructura del INEE comprenderá un área substantiva compuesta por tres subsistemas, al frente de cada uno de los cuales habrá un director, con el personal académico de planta y el personal externo que colabore por convenio o por honorarios. El personal académico de planta de cada área trabajará en forma colegiada.

De lo dicho antes se desprende que los subsistemas tendrán dimensiones diferentes, siendo el de pruebas de aprendizaje el mayor; el trabajo de elabo-

ración y aplicación de pruebas y las especificidades de cada variante de estas últimas (por ejemplo: pruebas de Matemáticas vs. de Lectoescritura; de conocimientos vs. habilidades; de respuesta estructurada o de ejecución; construidas con referencia a norma o criterios, etcétera.) harán necesario que haya responsables y grupos de trabajo diferentes para cada prueba o grupos asimilables de pruebas.

Además de un área de apoyo administrativo y la contraloría de ley, el INEE deberá tener otras áreas para apoyar el trabajo de los tres subsistemas básicos:

- Un área técnica, con personal académico que apoye a los subsistemas en los aspectos psicométricos y relacionados con la tarea de investigación.
- Un área que coordine las relaciones con las entidades federativas, la reproducción de instrumentos y el proceso de aplicación con su compleja logística.
- Un área editorial y de difusión, responsable de la publicación de los resultados de las evaluaciones y de otros materiales de apoyo.
- Un área de capacitación, para el personal propio y el subcontrado, así como de usuarios de la evaluación y evaluadores educativos de diferentes dependencias de educación básica.
- Un área de informática, que apoye los procedimientos de análisis y almacenamiento de la información, así como de la difusión electrónica de sus resultados.
- Un área para la coordinación de las actividades internacionales.
- Un área de biblioteca.

Aunque la investigación sea una función que el INEE desarrollará, no deberá haber un área especial para ello; serán los profesionales de cada subsistema, académicos del mayor nivel posible, quienes deberán hacer investigación como tarea concomitante a la de desarrollar nuevos instrumentos, y a la de analizar los resultados obtenidos mediante su aplicación.

La relativa complejidad del organigrama que se desprende de lo anterior, no implica algo excesivo; el personal contratado por honorarios debería representar una proporción alta del necesario para labores de apoyo, así como para algunas actividades de elaboración de pruebas (por ejemplo: redacción de reactivos) y para las de aplicación.



El personal de base será académico en su mayor parte, en el número indispensable para las tareas permanentes.

Estructura de gobierno

Buscando asegurar un funcionamiento eficiente y de alto nivel técnico, y así evitar los problemas que suelen derivarse de la politización y la influencia de intereses ajenos a la misión de las instituciones en la toma de decisiones, e independientemente de la personalidad jurídica que se decida dar al nuevo organismo, se propone una estructura de gobierno integrada por tres cuerpos colegiados, con diferente composición y funciones: la Junta Directiva y los Consejos Técnico y Consultivo.

- La Junta Directiva, máxima autoridad del Instituto. Dentro del límite de 15 miembros, si el INEE fue creado por decreto presidencial, o rebasándolo si se creó por ley, se buscó que su integración fuera la más adecuada para la credibilidad del nuevo organismo, en apoyo de su función de apoyo a la rendición de cuentas.

La Junta incluyó a funcionarios del sistema educativo, a cuyo servicio está el INEE, pero debe haber también representantes de sectores de la sociedad como los maestros, padres de familia y el sector empresarial, así como de instituciones académicas y organismos de investigación educativa. Para reforzar la independencia y credibilidad del INEE, y propuso que el papel de la Junta Directiva se limitase a las decisiones administrativas y que las de carácter técnico se asignaran a otro órgano, formado por especialistas.

La Junta Directiva aprobó la estructura orgánica del Instituto y sus reglamentaciones internas, fijaron los lineamientos que guiarán su operación administrativa, aprobó sus programas de trabajo anuales y los informes respectivos, y designó a los funcionarios de primero y segundo nivel. También designó a los integrantes del Consejo Técnico.

- El Consejo Técnico está constituido por 16 expertos reconocidos en evaluación o investigación educativa, mexicanos y extranjeros, designados por su mérito individual; se organizó en cuatro subgrupos: uno para aspectos generales y

uno más para cada uno de los tres subsistemas básicos del INEE. El Consejo Técnico fue presidido por uno de los integrantes del subgrupo de aspectos generales.

La función del Consejo Técnico es la de asesorar a la Junta Directiva y a la Dirección del Instituto en el desarrollo de sus funciones, apoyando la elaboración de los programas de trabajo, dando orientaciones técnicas para su desarrollo y evaluando sus resultados. Sancionará en definitiva las decisiones de carácter técnico y su opinión será consultada por la Junta Directiva para tomar las máximas decisiones que le corresponden. Podrá proponer personas calificadas para la designación de funcionarios. Designará también jurados para realizar los concursos para la selección del personal académico del Instituto.

El Consejo Técnico se define, como la instancia más importante para que el funcionamiento cotidiano del INEE sea el deseable; por ello la designación de sus primeros miembros fue crucial, y se buscó a las personas más calificadas para la tarea que se les encomendará. Para la designación de los futuros miembros de esta instancia, que substituyan a los anteriores en ciertos plazos, se puede pensar en que se haga a partir de propuestas del Consejo Técnico.

- El Consejo Consultivo, integrado por el responsable de evaluación de la Secretaría de Educación Pública y los de las 32 entidades federativas, constituye una instancia representativa de los usuarios directos de los productos de la actividad del Instituto, cuya opinión deberá ser tomada en cuenta para el desarrollo de dicha actividad.

4.5. Financiamiento

El financiamiento del organismo debe asegurar su independencia y su estabilidad, gracias a la disposición de recursos suficientes para el desarrollo de sus funciones, en una forma que permita la planeación a mediano plazo, máxime teniendo en cuenta que los tiempos requeridos para la consolidación de un sistema de evaluación educativa en gran escala son muy largos.

La operación de un buen sistema de evaluación implica costos importantes, que deberá cubrir el financiamiento público; dichos costos, por elevados



que puedan parecer, siempre representarán una fracción muy pequeña del costo del Sistema Educativo Nacional, y el impacto potencial sobre la calidad del sistema de ese gasto es tal que se justifica plenamente, siempre y cuando el trabajo y los productos derivados sean también de calidad.

Teniendo en cuenta la importancia de la evaluación educativa como tarea permanente, el INEE deberá contar de manera regular con los recursos presupuestales necesarios para el adecuado desarrollo de sus funciones. En un futuro razonable puede preverse que, además, complementen sus recursos con ingresos derivados de la venta de servicios a los sistemas educativos de las entidades federativas que lo soliciten, a escuelas particulares, e incluso a sistemas o subsistemas, educativos de otros países.

Los recursos públicos deberán incluir lo necesario para la constitución de un patrimonio formado por inmuebles propiedad del Instituto, y así evitar el pago de arrendamientos costosos; es deseable que se cuente también con un fondo de recursos financieros fideicomitidos, cuyos productos se utilicen para el mantenimiento y la renovación de equipos y bienes muebles e inmuebles.

El INEE deberá obtener apoyos específicos, para la realización de proyectos particulares, tanto por parte de la SEP (por ejemplo, para participar en proyectos internacionales, para desarrollar nuevas pruebas en áreas no cubiertas del currículo o de orientación novedosa, etcétera.) como de otras fuentes como el Conacyt.

La venta de servicios a usuarios diferentes de la SEP no deberá hacerse en condiciones monopólicas; por el contrario, el INEE deberá competir con otros proveedores de servicios, nacionales e internacionales, existente en el mercado. Deberán exigirse los mismos niveles de calidad a todos los proveedores, con base en los estándares propios del campo, teniendo en cuenta la importancia del impacto potencial de las evaluaciones. Los presupuestos del INEE deberán considerar los siguientes renglones:

Recursos humanos

Es claro que se trata del elemento central. Se buscó que el INEE fuera una institución relativamente pequeña, la cual cuente sólo con el personal de planta

necesario para el adecuado cumplimiento de sus funciones, lo que incluye, además de integrantes de los órganos colegiados de gobierno y de consulta, y de directivos capaces como:

- Un conjunto de personal profesional, de alto nivel académico, con estatus de investigador, que pueda formar parte del Sistema Nacional de Investigadores y sea suficiente para las tareas más complejas, como el diseño de modelos e instrumentos de evaluación que cumplan con los estándares de calidad pertinentes; el manejo eficiente de los procesos de aplicación de instrumentos y obtención de la información primaria; el procesamiento y análisis de la información, hasta la obtención de resultados y la construcción de los indicadores correspondientes; la capacitación de usuarios, desarrollando la capacidad técnica en los estados; y la difusión de resultados, fomentando la cultura de evaluación.
- Personal profesional del nivel requerido para realizar funciones de apoyo delicadas, como las que implica el procesamiento computarizado de datos y el almacenamiento de los mismos; la organización del trabajo con grupos de redactores o jueces; la logística de la aplicación de los instrumentos; el manejo del centro de información y los aspectos administrativos.
- Un mínimo indispensable de personal operativo de planta para tareas administrativas y de apoyo menos complejas, buscando que el mayor número de actividades de esta naturaleza se realice con apoyo de personal externo contratado para el caso.
- La cantidad necesaria de personal externo con perfil académico, que se contrate por honorarios para tareas como la redacción y la revisión de ítems o procesos de jueceo.
- Cierta cantidad de personal académico, de alto nivel, perteneciente a otras instituciones del país o el extranjero, con las que se hayan celebrado convenios de colaboración, o a quienes se contrate por honorarios para tareas específicas, sobre todo de capacitación. Con la idea de aprovechar la capacidad instalada en algunas instituciones académicas podrá pensarse también en la subcontratación de procesos completos de elaboración o aplicación, de algunos instrumen-



tos, cuidando siempre que se respeten los estándares de calidad más rigurosos.

Recursos materiales

- Serán los suficientes para el pago de los servicios personales que implica el inciso anterior.
- Los de gasto corriente para cubrir los conceptos de materiales, gastos de viaje, viáticos y otros, como los necesarios para reproducción de pruebas y materiales que requieran los modelos de

evaluación, y los que implica las actividades de capacitación y difusión.

- Los de inversión inicial y periódica para tener instalaciones suficientes y decorosas, propias o arrendadas; el mobiliario correspondiente; equipo de cómputo de punta para el manejo de grandes volúmenes de datos, así como para el trabajo del grupo académico; recursos bibliohemerográficos (obras de referencia, monográficas y publicaciones periódicas en papel o en línea) los cuales apoyen un trabajo de alto nivel académico.



5. Perspectivas de desarrollo

La constitución y puesta en marcha del INEE requirió de muchas acciones particulares, que serán las que pongan los cimientos de la nueva institución, por lo que su adecuada realización fue fundamental para el futuro; esto, a la vez, supone una perspectiva de mediano y largo plazos, que constituye una visión orientadora de esos primeros pasos.

Por ello en este apartado se presenta una perspectiva del desarrollo del INEE a mediano y largo plazos, con una visión en tres horizontes: al año 2002; al final del sexenio en el año 2006 y a un plazo mayor, en el 2012.

En el año 2002: El INEE terminó la etapa de arranque; con un estatus legal definido y adecuado a su misión, así como los elementos de infraestructura y de recursos materiales necesarios; se incorporó un grupo inicial de personal académico de base, con la calificación adecuada. Se envió a prepararse en posgrados en el país y el extranjero a un grupo de personas jóvenes con potencial para el trabajo de investigación evaluativo, y se tuvieron convenios de colaboración con instituciones nacionales e internacionales, para aprovechar recursos humanos escasos. La etapa de transferencia de los elementos con la anterior DGE habrá terminado, y se iniciaron la aplicación de algunos instrumentos planeados con la nueva perspectiva de la institución; se contó con los primeros resultados y se construyeron los primeros indicadores del nuevo esquema. Se inició el trabajo de educación de usuarios y de fortalecimiento de una cultura de evaluación. Se mantuvieron relaciones estrechas con la SEP y las secretarías estatales de educación, para que el desarrollo

de la actividad del Instituto respondiese de la mejor manera posible, a las necesidades del Sistema Educativo Nacional.

Al año 2006: El INEE alcanzó un estado de consolidación básica; desarrolló una gama de instrumentos convencionales de calidad, y comenzó la introducción de modelos avanzados de evaluación; contó con un grupo básico de personal académico de alto nivel y se integraron algunas personas enviadas a formarse en posgrados especializados; contó con una infraestructura suficiente para cumplir con sus funciones y seguir desarrollándose; su trabajo logró un desarrollo significativo de la cultura de la evaluación, y sus productos cubrieron ya una parte importante del universo de necesidades, sirviendo de manera significativa para la toma de decisiones.

Al año 2012: El INEE se habrá consolidado plenamente; sus instrumentos de evaluación habrán incorporado los mejores avances técnicos; participará activamente en el trabajo internacional de evaluación y, con un personal de alto nivel académico y una infraestructura técnica y bibliohemerográfica excelente, el INEE se habrá transformado en un centro de excelencia de nivel internacional.

La cultura de la evaluación se habrá difundido en los diferentes medios de la sociedad, en parte gracias a la acción del INEE, de manera que sus aportaciones evaluativas serán entendidas y valoradas por los actores involucrados y la sociedad en general, constituyendo elementos valiosos para el diseño de políticas educativas.



6. Conclusión

Un destacado investigador de los sistemas de evaluación que se están desarrollando en diversos países del mundo, después de comparar los casos de España, Francia, Inglaterra, Suecia y Portugal, encontró que en todos los casos se han establecido instituciones, las cuales *dependen de los más altos niveles de los ministerios y gozan, en general, de una gran independencia científica y metodológica* (Mateo, 2000: 244), y señala la necesidad de que cualquier organismo de esta naturaleza cumpla con dos principios básicos:

Un primer principio fundamental es conseguir una elevada independencia política a la hora de diseñar un modelo de evaluación del sistema educativo y de llevarlo a término... tanto en su práctica diaria, como en la percepción que la sociedad tiene de este organismo. El segundo principio es el de que los organismos encargados de evaluar han de tener un elevado nivel técnico-pedagógico, para garantizar la calidad de los procesos y la validez, fiabilidad y credibilidad de sus resultados. (Mateo, 2000: 244)

Es fácil apreciar que el Instituto Nacional para la Evaluación de la Educación, tal como se contempló en este proyecto, se apegó a esos dos principios. En síntesis, puede decirse que el nuevo organismo, tal como se propuso en el proyecto:

- Es un ente público, cuya personalidad jurídica, estructura orgánica y de gobierno, su personal y principios de operación le dan un amplio margen de autonomía técnica dentro del margen que permite la legislación, asegurando su credibilidad.
- Es una institución de carácter académico y técnico, con un trabajo de alto nivel centrado en la medición del aprendizaje de los alumnos del sistema nacional de educación; a partir de lo cual, y de actividades complementarias de evaluación de escuelas y de investigación de factores asociados con el rendimiento escolar, el INEE podrá

construir un sistema indicadores de calidad que cubran de manera completa y consistente la realidad educativa, como elemento de primera importancia para el diseño e implementación de políticas que hagan posibles procesos de mejoramiento sostenido de la calidad de las escuelas mexicanas.

- Con plena conciencia de la estrecha relación que existe entre currículo y evaluación, tiene como referente básico, para su trabajo, los planes y programas de estudio oficiales, buscando que los instrumentos de evaluación privilegien contenidos fundamentales y manejen una amplia gama de niveles de dominio, haciendo posible la realización de evaluaciones longitudinales en el mediano y largo plazos, más allá de los cambios curriculares que tengan lugar; debe cuidar que la evaluación ayude a centrar la atención de los maestros en los objetivos fundamentales de mayor nivel, apoyando a la autoridad educativa para evitar que se enseñe en función del examen o que pruebas de bajo nivel, desarrolladas por su mayor sencillez, se transformen en el currículo real.
- Sabiendo que la evaluación educativa es un esfuerzo cooperativo, el INEE trata de mantener relaciones productivas con universidades, centros de investigación y otras instituciones con propósitos similares a los propios, tanto en México como en el extranjero, buscando el reforzarse mutuamente. y formar una masa crítica de personal de alta calificación en este campo especializado del quehacer académico.
- Por último, entendiendo, que la evaluación debe servir al mejoramiento de la educación, para lo cual se requiere de la participación de toda la sociedad, realiza su labor atendiendo a los lineamientos de las orientaciones generales conteni-



das en este documento, prestando atención a la dimensión de equidad y esforzándose siempre, junto con las autoridades educativas, por hacer comprender a los demás actores sociales, en especial los maestros y los padres de familia, el sentido constructivo de la evaluación.

El último punto de los enunciados debe ampliarse, para concluir el documento, para atender una vez más las preocupaciones, esperanzas y temores de los diversos sectores de la sociedad en relación con la evaluación; en especial de maestros, padres de familia y el sector empresarial, además de las autoridades educativas.

Es importante que todos los actores tengan una comprensión cabal de lo que puede aportar la evaluación y qué no es razonable esperar de ella, para evitar tanto falsas expectativas como temores infundados.

La evaluación, como se mencionó, no produce por sí misma mejora ni deterioro; éstos dependen de las medidas que se tomen después. La evaluación es, se dijo, una de las estrategias particulares que debe comprender una *estrategia global de calidad*. El INEE será uno de los elementos del *Sistema Nacional de Evaluación Educativa*, entre otros.

Por ello conviene reiterar la distinción de los papeles de las autoridades educativas y los demás actores del sistema educativo, frente al que corresponde

a la evaluación y, en concreto, a un organismo como el INEE.

Se mostró ya la conveniencia de que las tareas de medición para la evaluación educativa se manejen con un grado suficiente de autonomía, con respecto a las autoridades y demás actores del sistema educativo, para asegurar la imparcialidad y objetividad máxima de los resultados, sin que lo anterior signifique aislamiento, para que la evaluación esté claramente articulada con las demás estrategias de mejoramiento de la calidad.

Debe señalarse: tampoco es adecuado que las instancias responsables de la evaluación pudieran tomar decisiones, o substituir a las autoridades educativas; si esto ocurriera simplemente se habría desplazado la autoridad, entonces la coincidencia de juez y parte en la evaluación persistiría.

La única solución será una donde los papeles se distingan con claridad; y los ámbitos de la autoridad y las instancias de evaluación se reconozcan y respeten; asimismo en cada parte deje que la otra, cumpla con su cometido y cada una, aporte lo que debe para el mejoramiento de la calidad.

Estas condiciones se cumplieron: el Instituto Nacional para la Evaluación de la Educación es un elemento valioso para el mejoramiento del Sistema Educativo Nacional.

Referencias

- Velázquez Castañeda, Víctor M. (s/f). *Sistema Nacional de Evaluación Educativa. Propuesta general*. México, SEP. Documento mecanográfico.

ANEXO I

REVISIÓN DE LITERATURA SOBRE

Introducción

I. Subsistema de indicadores. los indicadores en el mundo

- i.1. El proyecto de indicadores de la OCDE
- i.2. Los indicadores educativos en los Estados Unidos
- i.3. El sistema de indicadores de Francia
- i.4. Los indicadores educativos en otros países europeos
- i.5. Indicadores educativos en países africanos
- i.6. Un sistema latinoamericano
- i.7. Propuesta de indicadores educativos para el Estado de Aguascalientes

2. Subsistema de pruebas de aprendizaje

- 2.1. Los antecedentes de la medición educativa
- 2.2. Desarrollos teóricos y técnicos en el campo de las pruebas
- 2.3. Experiencias de evaluación en gran escala en otros países
- 2.4. La evaluación y la medición educativa en México
- 2.5. Conclusión

3. Subsistema de evaluación de escuelas

- 3.1. La escuela y la calidad
- 3.2. La evaluación de escuelas y los indicadores de la OCDE
- 3.3. El sistema inglés GRIDS
- 3.4. Los sistemas español y francés
- 3.5. Un sistema norteamericano
- 3.6. Los sistemas de inspección o supervisión
- 3.7. Conclusión

4. Conclusión general

Cuadro sinóptico

Bibliografía



Introducción

En este texto se presenta el resultado de una extensa revisión bibliográfica sobre los sistemas de evaluación de numerosos sistemas educativos, como referente básico para el diseño del Instituto Nacional para la Evaluación de la Educación (INEE). Se desarrollaron las experiencias relativas a los tres subsistemas que se propusieron como componentes básicos del INEE, el de indicadores educativos, el de pruebas para evaluar los aprendizajes y el de evaluación de escuelas.

En México, desde los años setenta la SEP comenzó a realizar actividades que llevaron a la creación de la Dirección General de Evaluación, la cual desde entonces ha aplicado pruebas de aprendizaje a muestras nacionales de alumnos de primaria. Sin embargo, esas actividades, no se consolidaron pronto; los resultados no se difundían y no se aprovechaban para la toma de decisiones. En el sexenio 1995-2000 la situación comenzó a modificarse: algunas entidades de la República comenzaron a desarrollar sistemas de evaluación propios y la SEP comenzó a desarrollar un Sistema Nacional de Evaluación Educativa, y dio pasos importantes en esa dirección, con lo que se dio un desarrollo sin precedentes en la evaluación educativa. Se han hecho esfuerzos para establecer las bases de una cultura de la evaluación, pero, parece claro, que los avances no son suficientes, y se debe seguir impulsándolo.

A nivel internacional, los fundamentos de las técnicas de medición se remontan al siglo XVIII, y su aplicación en las ciencias del hombre, comienza a darse en la segunda mitad del XIX. Para mediados del si-

glo XX la *teoría clásica de las pruebas* estaba plenamente desarrollada y a mediados del mismo, comenzaron a surgir los planteamientos de la *teoría de respuesta al ítem*, la *de la generalizabilidad* y otros. Al comenzar el siglo XXI puede hablarse de una nueva generación de pruebas, muy distintas a las de 1950. Contra una opinión que no es raro escuchar, las pruebas de tipo objetivo, lejos de tender a desaparecer se utilizan en un número creciente de países, al grado que es posible afirmar que ninguno importante está quedando al margen de la tendencia a utilizar este tipo de instrumentos. El número de experiencias nacionales e internacionales de evaluación educativa, en especial en el ámbito de la educación básica es considerable, y su revisión constituye un paso obligado como insumo para el diseño de un organismo dedicado precisamente a tal actividad, como es el del Instituto Nacional para la Evaluación de la Educación.

El resultado de una revisión así es el contenido de este documento, el cual se basó en tres textos anteriores, actualizados con información más reciente. Las referencias de los textos mencionados son las siguientes:

Martínez Rizo, Felipe (1996). *La calidad de la educación en Aguascalientes. Diseño de un sistema de monitoreo*. Aguascalientes, UAA-IEA, pp. 55-117 y 141-150.

Martínez Rizo, Felipe (1999). *El uso de pruebas en gran escala en educación. Antecedentes y problemática en el mundo y en México*. Mecanograma no publicado.

Martínez Rizo, Felipe (2001). Evaluación educativa y prebas estandarizadas. Elementos para enriquecer el debate. *Revista de la Educación Superior*. Vol. XXX, N° 120.

I. Subsistema de indicadores. Los indicadores en el mundo

En forma relacionada con la anterior, pero con algunas diferencias, desde los años sesenta se trató de identificar conceptos particulares precisos que permitieran concretar y valorar otros tan amplios e imprecisos como *desarrollo integral* o *bienestar social*. El concepto de Producto Interno Bruto (PIB), tan caro a los economistas, fue durante mucho tiempo el único criterio para apreciar el grado de desarrollo de un

Según Scheerens los indicadores son estadísticas globales, de baja inferencia, definidas a un elevado nivel de agregación (1994:17). Sin embargo, como veremos también, pueden encontrarse indicadores que son todo lo contrario: estadísticas específicas y de alta inferencia derivadas de encuestas de opinión o de investigaciones. Según Wyatt la definición más aceptada en la actualidad, se deriva de la de Oakes, según la cual un indicador debe ofrecer, por lo menos, uno de los siguientes tipos de información, sobre:

- Los logros de un sistema educativo en lo relativo a la obtención de determinados resultados; el indicador se liga, por tanto, a los objetivos, y constituye un punto de referencia para medir los avances (*benchmark*).
- La investigación ha mostrado algunas características relacionanadas con ciertos resultados; el indicador tiene así valor predictivo, porque su modificación traerá consigo otros cambios.
- Rasgos centrales del sistema (por ejemplo: insumos), o problemas, para entender cómo funciona.



- Aspectos relevantes para la política educativa, los cuales permitan apoyar la toma de decisiones.

Oakes, además, señala que los indicadores deben tener las siguientes características técnicas:

- Medir aspectos que se encuentren, de alguna manera, en todo el sistema, y de esa forma puedan hacerse comparaciones.
- Medir aspectos duraderos del sistema, de suerte que puedan analizarse tendencias y cambios en el tiempo.
- Ser fácilmente inteligibles para una audiencia amplia.
- Tener factibilidad, teniendo en cuenta el tiempo, costo y capacidad técnica requeridos para su obtención.
- Ser aceptables por su validez y confiabilidad (Wyatt 1994:105).

En este trabajo llamaremos *indicadores* a datos sintéticos, aprovechando las definiciones encontradas en la literatura, que resuman en forma numérica precisa, la situación actual de ciertos aspectos significativos de un sistema educativo (variables o conjuntos de variables). Los indicadores pueden organizarse en conjuntos sistemáticos, de acuerdo con alguna definición de calidad, como la que se desarrolla en otro documento, y pueden hacer referencia a algún estándar, o punto de referencia normativo, y entonces pueden servir como retroalimentación o llamada de atención para la toma de decisiones.

Si bien, desde tiempo atrás los sistemas educativos de muchos países manejaban estadísticas, éstas no conformaban sistemas integrados basados en una noción explícita de calidad, por lo que no constituían sistemas de indicadores en el sentido expresado. El desarrollo de sistemas de indicadores en este sentido es muy reciente: después de algunos intentos frustrados, como el que emprendió en 1973 la OCDE, la idea dio lugar a textos importantes a principios de los ochenta (Johnstone, 1981), y volvió a plantearse con más vigor a fines de la década, en diversos lugares (Véase Tuijnman Y Bottani, 1994:21).

1.1 El proyecto de indicadores de la OCDE

El más conocido y, sin duda, el más importante de los nuevos esfuerzos es el que encabeza una vez más

la OCDE. Su importancia es tanto mayor cuanto es claro no se limita a los trabajos del grupo especializado de la organización, sino que se prolonga en los países miembros, quienes desarrollan sus propios sistemas, más detallados por referirse a una escala menor y más enfocados a la problemática propia. No cabe duda, además, de que varios países no miembros de la OCDE, están desarrollando sus propios sistemas, en parte gracias al influjo de ella.

El proyecto INES (*Indicators of Education Systems*), comenzó a gestarse en una reunión convocada en noviembre de 1987 por la OCDE y los Estados Unidos, para discutir nuevos enfoques para el desarrollo de estadísticas educativas comparables (véase Tuijnman y Bottani, 1994:25). En mayo de 1988, el centro especializado de la organización para investigación educativa, el CERI (*Center for Educational Research and Innovation*) aprobó el proyecto, que pronto adquirió grandes dimensiones: además de un grupo técnico integrado por medio centenar de expertos de países pertenecientes, o no, a la organización y de los coordinadores nacionales de los países miembros, se formaron cuatro redes de expertos para trabajar temas específicos, cada red coordinada por un país:

- Una sobre los resultados educativos de los estudiantes (USA).
- Otra red sobre el destino de los estudiantes en el mercado laboral y la sociedad (Suecia).
- Una más sobre las características de las escuelas y los procesos que tienen lugar en su interior (Países Bajos).
- Y una cuarta red en relación con las opiniones, expectativas y actitudes de la población en relación con la educación (Reino Unido).

Además de múltiples productos preparatorios o parciales (véase CERI-INES 1991; IMHE 1993; Tuijnman Y Bottani 1994; CERI-INES 1995a; CERI-INES 1995b; CERI-INES 1995c). El resultado principal de estos trabajos es la publicación *Education at a Glance* (CERI-INES 1995) que contiene la información relativa a los indicadores del sistema de los países miembros de la OCDE. Publicada por primera vez en 1992, la obra tuvo dos ediciones más, en 93 y 95, y han aparecido versiones en varios idiomas. Cada edición fue ampliando el número de indicadores que contiene. La de 1995 se encontraron cerca de cubrir todos los previstos en el diseño del sistema.

A continuación se presenta la lista completa de indicadores incluyendo su clave, tal como se manejaban en 1995. Se puede notar que hay algunos espacios en la enumeración, los cuales se deben a que había algunos indicadores más: éstos estaban contemplados en el diseño del sistema pero que aún no se habían podido obtener para un número significativo de países.

Contexto demográfico

- ## Contexto social y económico

- ## Opiniones y expectativas sobre la educación

- | | |
|-----|--|
| C26 | Opiniones sobre las prácticas escolares prioritarias para el mejoramiento de las escuelas. |
| C27 | Opiniones sobre decisiones sobre la escuela que deberían tomarse en cada plantel. |

Gasto educativo

- ## Origen de los recursos educativos

- | | |
|-----|---|
| F11 | Porcentaje de recursos públicos y privados por nivel educativo. |
| F12 | Recursos públicos para educación por nivel de gobierno. Porcentaje de recursos centrales, regionales y locales en el gasto educativo público. |
| F13 | Porcentaje del gasto en educación en el total del gasto público. |

Por Participación de la población en la educación formal. Porcentaje de alumnos de tiempo completo o parcial en población de 5-29 años. Esperanza de escolarización de un niño de cinco años.

- | | |
|-----|---|
| Po2 | Educación de la primera infancia. Tasas netas de escolarización en preescolar y primaria. |
| Po3 | Educación media. Tasas netas de escolarización en tipos de educación media. |
| Po4 | Transición de la educación secundaria a la terciaria. Tasas de asistencia de alumnos de tiempo completo de 17 a 24, años por tipo de institución y año. |



- P05 Ingreso a educación terciaria. Porcentaje de alumnos de nuevo ingreso, a tiempo completo, a instituciones de educación terciaria, universitaria o no, en relación con el grupo de edad de ingreso teórico a ese nivel.
- P06 Participación en educación terciaria. Tasas netas de participación en educación de tipo universitario o no.
- P08 Educación continua y de adultos. Porcentaje de población ocupada de 25 a 64 años inscrita en programas de educación continua o adultos por nivel educativo y grupo de edad.

Tiempo dedicado a la enseñanza

- P11 Porcentaje dedicado a cada materia en el tiempo escolar total en la educación secundaria.
- P12 Total de horas de enseñanza anuales que deben seguir los alumnos de 9 y 14 años. Promedios y dispersión de las cifras.

Procesos escolares

- P21 Prácticas de agrupamiento de alumnos de nueve años en materias de lectura y lengua.

Recursos humanos

- P31 Personal docente y no docente del sector educativo en porcentaje del total de la fuerza de trabajo.
- P32 Proporción de estudiantes por docente por nivel educativo.
- P33 Número de horas de enseñanza activa que deben cubrir los maestros en los diversos niveles.
- P34 Total de años de educación, a partir del primero de primaria, requeridos para poder trabajar como maestro de cada nivel.
- P35 Salarios inicial y máximo de profesores de educación pública de educación básica.
- P36 Distribución por edad y género de los maestros de los diversos niveles.

Actividades de investigación y desarrollo en educación (I&DE)

- P41 Personal equivalente de tiempo completo (ETC) dedicado a I&DE. Número absoluto y porcentaje del total de investigadores y del subtotal en instituciones de educación superior.

- P42 Gasto I&DE. Total de gasto I&DE público y privado y porcentaje del gasto educativo total y del total del gasto en I&D.

Indicadores de resultados educativos

Resultados de alumnos

- R04 Nivel de aprendizaje de lectura de alumnos de 9-14 años, según pruebas estandarizadas.
- R05 Frecuencia de lectura reportada por alumnos de 14 años, en relación con resultados en pruebas objetivas.

Resultados del sistema

- R11 Porcentaje de graduación en media superior en relación con grupo de edad por tipo de programa.
- R12 Porcentaje de graduación en primer nivel universitario en relación con grupo de edad, por género.
- R14 Número y porcentaje de primeros universitarios por área del conocimiento respecto al total.
- R15 Graduados universitarios en áreas científicas y tecnológicas (ingenierías) sobre total y porcentaje de graduados en ellas por 100 mil personas de 25-34 años de edad en la PEA.

Resultados en el mercado laboral

- R21 Desempleo-educación. Tasas de desempleo por nivel educativo de la población de 25 a 64, y de 25 a 34 años, por género.
- R22 Ingresos-educación. Proporción de ingresos anuales promedio, por nivel educativo, en relación con ingresos medios de los egresados de educación media superior en una población de 25-64 años por género.
- R23 Logros educativos de los trabajadores. Índice de años de estudio de trabajadores 35-54 años, en ciertas ocupaciones.
- R24 Tasas de desempleo de jóvenes que abandonan el sistema educativo en diversos niveles.

Los apartados del sistema de indicadores de la OCDE en su versión de 1995 y sus apartados pueden verse más en forma sintética en la siguiente tabla:

Sobre el contexto de la educación	
Contexto demográfico	3
Contexto social y económico	3
Opiniones y expectativas sobre la educación	7
Sobre costos, recursos y procesos escolares	
Gasto educativo	5
Origen de los recursos educativos	3
Tasas de participación en la educación	7
Tiempo dedicado a la enseñanza	2
Procesos escolares	1
Recursos humanos	6
Actividades de investigación y desarrollo en educación	2
Sobre resultados educativos	
Resultados de alumnos	2
Resultados del sistema	4
Resultados en el mercado laboral	4
Total	49

- Llama la atención la amplitud de la lista de indicadores; se está muy lejos del simple *conteo de insumos* de las estadísticas tradicionales.
- Puede notarse que se trata de indicadores concebidos para países relativamente desarrollados. Algunos no parecen muy viables en países con menor desarrollo y, en cambio, en éstos harían falta otros.
- El grado de detalle de algunos indicadores de gasto, empleo, ingresos, etcétera, sólo es posible si los países cuentan con buenos sistemas de estadísticas económicas y laborales.
- En otros casos, como los de resultados escolares y los de opiniones y expectativas, los indicadores son posibles si existen otros sistemas para el de-

- El apartado de procesos escolares sigue siendo muy pobre, pese al esfuerzo por enriquecer el listado, lo que es explicable por la complejidad de esa dimensión de la realidad educativa y hace pensar en la necesidad de contar con enfoques complementarios.

Indicadores manejados en *Education at a glance 2000*

A 1	Tamaño relativo de la población en edad escolar: • Porcentaje en el total de la población de las personas del grupo de edad de 5 a 14 años (población en edad de asistir a la educación básica); • Porcentaje del grupo de 15 a 19 años (en edad de asistir a la educación media-superior); • Porcentaje del grupo de 20 a 29 años (en edad de asistir a la educación superior); • cambios en la proporción de dichos grupos de edad en años anteriores; • cambios previstos en los años siguientes.
A 2	Logro educativo de la población adulta (desagregado por género): • Porcentaje de la población 25-64 años que alcanzó determinado nivel educativo;



- Porcentaje de la población 25-34 y 55-64 con al menos educación media superior o educación superior que permite acceder a programas avanzados de posgrado;
- Porcentaje de la población con al menos nivel 3 en escala de alfabetización del IALS, según grupo de edad.
- Estos indicadores corresponden a los de contexto demográfico en el esquema de 1995, en particular a los indicadores C 01, C 02, y C 03.

B. Recursos financieros y humanos invertidos en educación

- B 1 Gasto en educación en relación con el PIB: gasto en instituciones educativas (IE) por origen de los fondos, distinguiendo, pagos privados a las IE; gasto público directo en las IE, y subsidios públicos a las familias para pagos a IE, no incluyendo subsidios para sostenimiento. Se calcula tanto globalmente como por nivel, distinguiendo educación primaria, secundaria y postsecundaria, no terciaria, y educación terciaria. Gasto público en educación como proporción del total del gasto público. Se calculan también índices de cambio.
- B 2 Proporciones relativas de gastos públicos y privados en educación, globales y por nivel educativo. Porcentajes de estudiantes inscritos en instituciones de educación superior que cobran algo por concepto de matrícula o colegiatura. Índices de cambio.
- B 3 Subsidios públicos a las familias y entidades privadas, incluyendo reducciones de impuestos, como porcentaje del gasto público total en educación terciaria.
- B 4 Gasto educativo por estudiante. Gasto anual en IE públicas o privadas, por nivel. Gasto educativo en relación con el PIB *per cápita*. Relación entre el gasto por alumno en el nivel primario y en otros niveles educativos.
- B 5 Gasto educativo en inversión *vs.* gasto corriente y, dentro de este último salarios (de

maestros, de otro tipo de personal y totales) y otros rubros de gasto corriente.

- B 6 Gasto público en educación por nivel de gobierno quien lo aporta y lo ejerce, distinguiendo los niveles central, regional y local. A partir del cálculo de índices, se estimó la relación que puede haber entre la centralización o descentralización del financiamiento y la de la toma de decisiones en varios aspectos.

- B 7 Ratios de alumnos/docentes, en términos de maestros equivalentes de tiempo completo, por nivel educativo.

Los indicadores equivalentes, o emparentados, en el esquema de 1995 fueron los identificados con las claves F 01 a F 05 y F 11 a F 13, así como el P 32. No se manejó el indicador P 31, personal docente, y no docente del sector educativo como porcentaje del total de la fuerza de trabajo.

C. Acceso a la educación, participación y progreso

En este rubro se presentan indicadores más finos que en 1995, manejando tasas netas *vs.* brutas, a partir de sistemas de información los cuales permitieran manejar cohortes reales, basados en conteos de individuos reales (head counts). Los indicadores de 2000 incluyen:

- C 1 Participación global en la educación:
- Expectativa de escolarización en las condiciones actuales.
 - Tasas netas de escolarización por año de edad y nivel educativo.
 - Características de la transición de la situación de escolarización de los 17 a los 20 años: tasas netas de escolarización por año de edad y nivel escolar.
- C 2 Participación en educación secundaria y terminación de la misma:
- Ratio de los graduados de educación media superior en relación con la población en la edad normal de graduación de ese nivel.
 - Distribución de la matrícula de media superior, por tipo de programa.
- C 3 Acceso la educación superior y participación

- | | |
|-----|--|
| | de horas de instrucción a la semana y el número de semanas al año. |
| D 5 | Ausentismo de alumnos. Porcentaje de alumnos de 8° que no asisten a la escuela en un día normal, por cualquier razón. |
| D 6 | Toma de decisiones sobre el currículo en educación media básica. |
| D 7 | Ratio alumnos/computadora en educación media básica, acceso a Internet, e importancia en el currículo y la práctica de actividades para el manejo de ese tipo de tecnología. |

Los indicadores del esquema de 1995 correspondientes al bloque anterior fueron los designados por las claves P 01 a P 08, y R 11 a R 15.

E. Resultados educativos individuales, sociales y laborales

- | | |
|-----|--|
| E 1 | Participación en la fuerza de trabajo por nivel de educación alcanzado, hombres y mujeres de 25 a 64 años. Empleo y desempleo por edad y nivel educativo. |
| E 2 | Educación y empleo de jóvenes de 20-24 años, distinguiendo los que estudian o no y los que trabajan o no, incluso los que participan en programas combinados estudio-trabajo. |
| E 3 | Situación específica de la población joven. Diferencia de la proporción de personas empleadas o no, parcial o totalmente, entre los jóvenes y la población total. |
| E 4 | Años esperados de educación y de trabajo de los jóvenes de 15 a 29 años. |
| E 5 | Ingreso y logro educativos. Ingresos relativos, en comparación con los que perciben quienes tienen educación media superior o postsecundaria no terciaria, de quienes tienen menos y más educación, en la población de 25 a 64 años, por género. |

Se relacionan con estos indicadores los designados con las claves C 11, C 12 y C 13, así como R 21 a R 24 en el esquema de 1995.



F. Desempeño de los alumnos

Gracias a las evaluaciones internacionales estandarizadas, en particular el TIMSS, los indicadores de la OCDE en 2000 incluyeron varios que se refieren al aprendizaje alcanzado por los alumnos en grados y materias clave. Las evaluaciones de PISA permitirán mantener y aumentar este tipo de indicadores, que en el pasado no era posible obtener.

- F 1 Desempeño en Matemáticas de alumnos de 4° y 8°.
- F 2 Diferencias en las actitudes hacia la ciencia de alumnos de 4° y 8°.
- F 3 Creencias de los alumnos de 4° y 8° en cuanto a su desempeño en Matemáticas.

Estos indicadores corresponden a los que en el esquema de 1995 se designaban con las claves R 04 y R 05, y substituyeron a los que se basaban en opiniones y expectativas de los padres de familia sobre lo que se enseña en las escuelas (C 21 a C 27). No se manejaron los indicadores relativos a actividades de investigación y desarrollo.

1.2 Los indicadores educativos en los Estados Unidos

En Estados Unidos, con su gran sistema educativo, el interés por la calidad de la educación y su evaluación, tiene antecedentes muy antiguos, pero se ha manifestado con especial fuerza en las últimas décadas, al grado de que las deficiencias de las escuelas llegaron a ser consideradas motivo de preocupación en relación con la seguridad nacional, tan cara a los gobiernos republicanos, como en la administración del Presidente Reagan. El título del más famoso de una serie de reportes especializados de la primera mitad de los años ochenta es sintomático: *A Nation at Risk* (1983).

Con todas las salvedades que se quiera hacer, aceptando que una visión catastrofista carece de fundamento es, seguramente, innegable que había motivos reales para preocuparse por la calidad de la educación en los Estados Unidos hace 15 años, y que siguió habiéndolos durante los ochenta. De la aparición de *A Nation at Risk*, hasta las famosas seis metas para la educación del plan América 2000 adoptadas por el Presidente Bush y los gobernadores de los 50 estados en su histórica reunión de 1989, las manifestaciones

de esta inquietud son numerosas. Paralelamente, persistió y se fortaleció el interés por contar con sistemas de información y evaluación confiables, que permitieran fundamentar las afirmaciones hechas sobre la calidad de las escuelas, evitando extremos basados en presupuestos más que en evidencias reales. Por su naturaleza federal, todavía en la segunda mitad de los años ochenta los Estados Unidos, en donde se han desarrollado con más fuerza los trabajos de investigación educativa en general, y de investigación evaluativa en particular, carecían de un sistema nacional de estadísticas e indicadores educativos que permitiera comparar en forma homogénea a los sistemas y sus cincuenta entidades federativas. Muchos de éstos manejaban sistemas propios, algunas veces de buena calidad, pero heterogéneos entre sí. Kaagan y Coley informan que en esas fechas 23 de las 51 entidades federativas norteamericanas (incluyendo al Distrito de Columbia), contaban con sistemas de indicadores, en tanto que 48 aplicaban pruebas estatales de conocimientos en ciertos niveles educativos (1989).

Cuando se advirtió que prácticamente todos los estados manifestaron tener cifras mejores a la media nacional en ciertos indicadores, se cayó en la cuenta de la necesidad de contar con un sistema nacional que permitiera comparaciones interestatales válidas. Las *Wall Charts* que comenzó a publicar en 1985, no sin resistencias, el Centro Nacional de Estadísticas Educativas del Departamento Federal de Educación, con el impulso del Secretario Bell, fueron el primer esfuerzo en tal dirección (Bryk & Hermanson 1992:451). Posteriormente recientemente comenzó a aparecer la publicación anual, del mismo Centro Nacional de Estadísticas Educativas, *Condition of Education. The Digest for Educational Statistics*.

Pronto se cayó en la cuenta de que no bastaba homogeneizar las estadísticas existentes, que coincidían en su reducido enfoque centrado en el conteo de insumos. Se comenzó a ver claramente que, para tener elementos suficientes para juzgar la calidad de los sistemas educativos, en toda la amplitud de tan elocuente concepto, se necesitaban sistemas de información más completos y finos, los cuales atendieran no sólo a los insumos, sino también a los productos y a los procesos que tuvieron lugar en el interior de las escuelas, así como a las necesidades de su entorno social. En septiembre de 1991 se publicó el trabajo de

un grupo de especialistas (*Special Study Panel on Education Indicators*, SSPEI, 1991), contratados en 1989 por el Centro Nacional de Estadísticas Educativas para diseñar un sistema de indicadores para monitorear el *estado de salud del sistema educativo norteamericano*.

La extensión y calidad del grupo, así como el tiempo de que dispusieron para realizar la tarea encomendada, permitieron formar una idea de la conciencia existente sobre la importancia y la complejidad de la misma. De hecho, el resultado no es todavía un sistema completo de indicadores, sino apenas un excelente esquema del mismo. Los autores señalan desde el principio de su trabajo: la importancia de un sistema de indicadores radica en que no sólo refleja ciertas prioridades educativas sino también, en buena medida, las define.

Determinadas áreas del currículo son importantes, si se aplican pruebas y se formulan indicadores sólo sobre otras, serán estas últimas las que acaparen la atención de maestros y alumnos, quienes descuidarán las primeras: “si la nación está de acuerdo en que las Matemáticas y la Geografía son suficientemente importantes para tener su respectivo indicador, dice el documento, haciendo referencia a una de las metas educativas establecidas por la administración del Presidente Bush para el año 2000; en tanto que la Música o las lenguas extranjeras no, los educadores actuarán en consecuencia (SSPEI 1991, p. 7).

Por consiguiente, argumentan algunos autores, se debe evitar el error de diseñar un sistema de indicadores a partir de la información disponible, de lo inmediatamente factible, aunque se afirme que los aspectos no cubiertos por tal sistema también son importantes. Esta salvedad teórica pronto será olvidada en la práctica, y lo realmente importante será definido por lo que pueda *medirse* con los indicadores disponibles. Por el contrario, se debe diseñar el sistema de indicadores a partir de lo que en principio se considera importante, aunque para muchos aspectos no se cuente con datos adecuados y deban realizarse después todos los esfuerzos para contar con la información necesaria. El sistema resultante del trabajo del grupo plantea seis grandes conjuntos de indicadores, relacionados con otras tantas dimensiones del quehacer educativo, resultan fundamentales para una evaluación integral del mismo y son las siguientes:

- Qué y cuánto aprenden los estudiantes (*learning outcomes*).
- Qué tanto favorecen el aprendizaje las escuelas (*quality of educational institutions*).
- Qué antecedentes de los niños los preparan mejor para la escuela (*readiness for school*).
- Qué apoyo brindan las instituciones sociales al aprendizaje (*societal support for learning*).
- En qué medida la educación contribuye a la productividad económica (*education and economic productivity*).
- En qué medida la educación ofrece igualdad de oportunidades y contribuye a la equidad y al bienestar social (*educational equity*).

El trabajo aún no llega a un listado preciso de indicadores como el de la OCDE, pero incluye una serie de aspectos particulares de cada una de los seis grandes conjuntos o áreas mencionados:

Área I. Resultados del aprendizaje

Dominio de contenidos básicos: Inglés, Matemáticas, Ciencias Naturales, Ciencias Sociales y Humanidades, Música y otras artes, Lengua extranjera.

Habilidades integradoras, de razonamiento e interdisciplinariedad y para el trabajo: alfabetización científica y tecnológica, comprensión internacional, pluralismo, competencias para el trabajo.

Actitudes-disposiciones: tolerancia, autocontrol, responsabilidad, valoración de la asistencia a la escuela, compromiso con el aprendizaje, dedicación al oficio.

Área II. Calidad de las instituciones

Oportunidades de aprendizaje: tiempo efectivo de exposición en cada materia, naturaleza de oportunidades de aprendizaje (activas-pasivas, etcétera), asignación de grupos a maestros, integración curricular.

Maestros: nivel de ingreso de los que escogen la profesión, calidad de la formación inicial, competencia demostrada en el aula.

Condiciones de trabajo de los maestros: disponibilidad de recursos básicos en el aula, recursos de apoyo, capacidad de decisión sobre los contenidos, apoyo para el desarrollo profesional.



Clima institucional (las instituciones como lugares con objetivos y carácter): claridad de la misión, ambiente humano, orden y seguridad, prioridad de lo académico.

Recursos de las escuelas: edificios, bibliotecas, laboratorios, equipamiento, personal profesional de apoyo.

Área III: Preparación para la escuela

Situación de niños pequeños y sus familias: perfiles de niños de tres a seis años; salud, nutrición, peso al nacer, hogares de un solo cónyuge, cuidados prenatales y seguridad social.

Servicios: programas preescolares, características de las escuelas, currículo e instrucción.

Área IV: Apoyo de la sociedad para el aprendizaje

Apoyo familiar: actitudes de padres, participación en trabajo del hogar, responsabilidad por cuidados elementales, participación de los padres en las actividades de la escuela.

Apoyo de la comunidad: existencia de bibliotecas y de servicios en la comunidad para apoyo de materias particulares como ciencias, artes, etcétera.

Apoyo del ambiente cultural: actitudes cívicas y comportamiento electoral, hábitos de lectura, competencia de actividades como la televisión con lo educativo.

Apoyo económico: fuentes de los recursos escolares (federales, estatales, locales), distribución por objeto del gasto, recursos y gastos educativos de otras instituciones.

Área V: La educación y la productividad económica

Resultados formales: adquisición de competencias para el trabajo; persistencia, deserción, y eficiencia terminal; elección de diversos programas o materias económicamente importantes, como las de ciencia y tecnología; atención de alumnos destacados.

Consecuencias económicas de la educación: empleo, ingresos diferenciales según el nivel educativo alcanzado.

Apoyo de empresas a la educación: educación en lugares de trabajo; políticas de apoyo a la educación por parte de empresas.

La función de investigación de la educación superior: actividades académicas de investigación y desarrollo y su impacto.

Área VI: La equidad: recursos, características demográficas, estudiantes en situación de riesgo

Características demográficas: pobreza, grupos minoritarios, niños con discapacidades, alumnos que no dominan el Inglés, segregación racial, familias de un solo cónyuge.

Características de las instituciones educativas: tipo, control, ubicación, composición del alumnado.

Servicios educativos: expectativas diferenciales y agrupamiento inequitativo para estudiantes con desventajas; clima escolar; programas compensatorios especiales.

Comparando los puntos anteriores con los de la OCDE, es posible hacer algunos comentarios:

- Por una parte, es obvio el distinto grado de desarrollo de los dos trabajos.
- Por otra, se ven coincidencias de concepción, junto a diferencias de detalle.
- En tercer lugar, también se nota que este segundo trabajo se refirió al sistema educativo de un país desarrollado.
- Por último, y tal vez con mayor claridad que en el caso de la lista de la OCDE, se percibe la dificultad para obtener la información necesaria y poner en práctica un sistema de indicadores ambicioso.

Aún limitándonos a la primera de las seis dimensiones, la más convencional al pensar en evaluación educativa, que se refiere a qué y cuánto aprenden los estudiantes; diseñar un buen sistema de evaluación no es fácil. No se trata, como se hace con frecuencia cuando se abordan estas cuestiones por primera vez, de aplicar simplemente, en condiciones más o menos controladas, pruebas de rendimiento construidas sin una metodología rigurosa. Este procedimiento sencillo encierra cuestiones complejas de confiabilidad y validez, de las que se trató en el apartado relativo a los sistemas de pruebas estandarizadas. Por el momento continuaremos viendo sistemas de indicadores desarrollados en diversas partes.

1.3. El sistema de indicadores educativos de Francia

El sistema educativo comprende tres niveles: el primero incluye la educación preescolar (maternal, de dos a cinco años) y la educación primaria (cinco años a 11, designados como 1° a 5°). La educación secundaria comienza a los 12 años con el 6° que se designa tradicionalmente *6ème*. Los cuatro primeros años de la secundaria son obligatorios, con lo que la educación básica llega hasta 16 años. El 2° ciclo, no obligatorio, de secundaria, a partir de los 16, puede durar dos, tres o cuatro años y da acceso a diversos diplomas. La educación superior, en universidades y otras instituciones tiene tres ciclos: un diploma de dos años general (DEUG) o profesional; un 2° ciclo, profesional (Maestría y DEA); y el 3er. ciclo de doctorado.

El sistema educativo francés se ha distinguido por la solidez de su estructura, en el marco de un esquema altamente centralizado, marcado todavía por las reformas napoleónicas de principios del siglo XIX. En este contexto, el Ministerio de la Educación Nacional contó con eficientes mecanismos de recolección de datos, que se manifestaron en varias publicaciones: la serie *Reperes et références statistiques*, que se publica anualmente desde 1984; desde 1983, la publicación *Géographie de l'école*, que presenta de manera precisa la situación de las diferentes regiones de Francia; y especialmente *L'état de l'école*, que desde 1991 resume la situación del sistema educativo francés en una serie de treinta indicadores, o subconjuntos de indicadores.

En 1994 el conjunto se estructuraba siguiendo dos ejes: por una parte, los niveles del sistema educativo, distinguiendo indicadores relativos al conjunto (11), a la enseñanza primaria (4), a la secundaria (5), a la educación superior (6) y a la formación continua (4). El otro eje distinguió los indicadores de costos (insusos), los de actividades (procesos) y los de resultados (productos). Los ejes se combinan, de manera que hay indicadores de costos, actividades y resultados para el conjunto del sistema educativo y para cada uno de los niveles considerados. A continuación se enumeran los treinta indicadores de la versión 1994 del sistema. La sigla que precede al número de cada uno indica si se refieren a costos (C), actividades (A) o resultados (R).

Indicadores del conjunto del sistema

- C1. Gasto educativo. Global, público (por nivel de administración, central, regional), privado. Por habitante, por alumno y como porcentaje del PIB.
- C2. Gasto educativo del Ministerio de Educación Nacional, global y por alumno.
- C3. Personal del sistema educativo. Total, desglose por actividad (enseñantes y otros) por sexo y dependencia (Ministerio de Educación, otros). Porcentaje de la PEA en educación.
- A4. Esperanza de escolaridad de los alumnos que inician la escuela en un ciclo, por sexo.
- A5. Proporción de una generación que accede al nivel de bachillerato o de un diploma de educación profesional de nivel equivalente.
- A6. Ídem por región en cifras efectivas y comparadas con las esperables por composición social de la población.
- R7. Resultados de los conscriptos en cuatro pruebas de aptitud académica, por cohorte anual, según escolaridad.
- R8. Número de alumnos que salen del sistema educativo con diversos niveles de preparación.
- R9. Tasa de desempleo de jóvenes de menos de 25 años con o sin calificación, por sexo.
- R10. Diferencias de salario según nivel educativo entre personas de 25 a 29, y de 35 a 44 años.
- R11. Porcentaje de personas con o sin bachillerato o diploma superior, con una actividad de nivel medio o alto en el grupo de 25 a 29 años, controlando la profesión del padre.

Indicadores de la educación primaria

- C12. Gasto educativo en primaria.
- A13. Porcentaje de niños de dos a cinco años que están escolarizados.
- A14. Duración promedio que requiere un niño para terminar los cinco años de educación primaria.
- R15. Resultados obtenidos en pruebas de Francés y Matemáticas por los alumnos al entrar al primer año de educación secundaria (6°).



Indicadores de la secundaria

- C16. Gasto educativo en secundaria.
- A17. Inscripción en secundaria, total y por ciclo. Efectos del crecimiento demográfico y de los incrementos en las tasas de escolarización respectivas.
- A18. Duración anual teórica de la enseñanza recibida, por grupo de disciplinas.
- R19. Esperanza de obtener uno de los diplomas terminales de la educación secundaria para los miembros de las cohortes que ingresan al primer año de la misma (6°).
- R20. Tasas de transición entre grados de la secundaria de segundo ciclo y probabilidades de alcanzar el diploma de bachiller, por establecimiento. Resultados efectivos por establecimiento, comparados con lo esperable según composición del alumnado.

Indicadores de educación superior

- C21. Gasto en educación superior.
- A22. Porcentaje de egresados de bachillerato general o tecnológico que accede al año siguiente a diversas instituciones de educación superior.
- A23. Matrícula en educación superior. Efectos del crecimiento demográfico y los cambios en la escolarización. Tasas de escolarización de grupos por edad, de 18 a 25, y de 19 a 21.
- R24. Tasa de acceso al segundo ciclo universitario de los que entran al 1°, por tipo de diploma.
- R25. Porcentaje de alumnos que aprueban el segundo y tercer ciclo universitarios (maestría, DEA y doctorado).
- R26. Número de egresados de educación superior con diplomas de diferente nivel.

Indicadores de Formación Continua

- C27. Gasto en educación continua. Público y privado (empresas y familias).
- A28. Número de personas y horas en actividades de formación continua.
- A29. Ídem en establecimientos públicos de educación secundaria y superior.
- R30. Tasa de acceso a empleo de nivel superior tras haber realizado actividades de formación continua, por sexo, según diversas

duraciones de la formación y si ésta fue tomada por iniciativa personal o del contratista.

Es pertinente resaltar algunos puntos del sistema francés: por una parte, estamos nuevamente, como en el caso de la OCDE y a diferencia del trabajo americano reseñado, ante un sistema plenamente desarrollado, el cual tiene varios años de operación. Por otra, salta a la vista que se trata del sistema de un país desarrollado, con indicadores con excelentes sistemas de estadísticas laborales y económicas. Por último, algunos de los indicadores franceses contienen ideas novedosas que conviene destacar: una es la de utilizar los resultados de las pruebas que se aplican anualmente a los conscriptos (indicador R7); otro aspecto es el de comparar los datos reales de una región con los que podrían esperarse dada la composición de su población (A6, R20); una más es el uso del indicador relativo a la duración efectiva promedio para terminar la primaria (A14).

En la versión 2000 de *L'état de l'école* se mantienen treinta subconjuntos de indicadores, con la idea de que un sistema así no debe tener demasiados elementos, so pena de volverse pesado y poco accesible para los usuarios, pero aparecen indicadores nuevos y desaparecen otros. Debe subrayarse la aparición de bloques concernientes a los objetivos educativos más ambiciosos (como el que por el nivel mínimo de escolaridad alcanzado por todo ciudadano francés, sea el llamado nivel V y que un número creciente llegue al nivel siguiente, el IV, en los nuevos bloques 6 y 7), así como con una atención mayor al tema de equidad (Bloques 8, *la educación prioritaria*; 9, *los jóvenes en dificultad*; 11, *medio social y acceso al bachillerato y a la educación superior*).

1.4. Los indicadores educativos en otros países europeos

La Oficina de Estadísticas de las comunidades europeas emprendió un trabajo de elaboración de indicadores educativos complementario al de la OCDE, éste atiende temas de especial interés para esos países. Por ejemplo, se manejan los siguientes indicadores:

- Niños de familias migrantes en las escuelas
- Alumnos en instituciones que cobran cuotas en 1° y 2°
- Subsidios a alumnos de educación superior



11. Libros de texto y apoyo por alumno
12. Porcentaje de niños con pupitre
13. Proporción de alumnos por maestro
14. Porcentaje de maestros calificados
15. Porcentaje de maestros según niveles de experiencia
16. Egreso de las instituciones formadoras de maestros
17. Deserción de maestros
18. Maestros en áreas difíciles
19. Maestros que disponen de vivienda en la escuela
20. Proporción de asesores profesionales por maestro
21. Proporción de inspectores por maestro
22. Actividad del consejo escolar
23. Gasto educativo en relación con el PIB
24. Gasto educativo por nivel
25. Fuentes del gasto educativo
26. Gasto público en educación en relación con el total
27. Gasto por alumno
28. Gasto por alumno en relación con el PIB per cápita
29. Gasto por maestro
30. Gasto por maestro en relación con el PIB per cápita
31. Fuentes de los ingresos de los maestros

El análisis de ambas listas muestra coincidencias generales, de carácter conceptual, con las de los países más desarrollados. Al mismo tiempo aparecen grandes diferencias en el contenido detallado, reflejan, sin duda, el muy distinto nivel de desarrollo. El que un país maneje indicadores como la proporción de niños escolarizados que toman sus clases en aulas especialmente diseñadas para tal propósito, en contraposición a quienes lo hacen en iglesias o al aire libre, o como el porcentaje de niños que cuentan con pupitre (en 1° el porcentaje pasó del 8 al 14 porcentaje entre 1985 y 1990). Está claro que estos dos sistemas no suponen la existencia previa de sistemas de estadísticas nacionales sofisticados, sino que pueden desarrollarse mediante un esfuerzo razonable del sistema educativo.

1.6. Los indicadores educativos en América Latina

El Laboratorio Latinoamericano de Evaluación de la Calidad Educativa propone un sistema que concreta su modelo de calidad. El sistema combina seis *áreas de variables incidentes en el aprendizaje escolar* (política social-educativa, familia, escuela, alumno, currículo y

profesor), con las dimensiones de insumos y procesos. Cuenta, además, con indicadores de productos.

Indicadores de insumos

Política social-educativa:

- Gasto en educación del país y por alumno
- Escolaridad obligatoria

Familia del estudiante:

- Composición del grupo familiar y nivel ocupacional-educacional de los padres
- Equipamiento del hogar. Bienes y recursos culturales

La escuela:

- Ubicación geográfica (rural, urbana)
- Dependencia administrativa (particular, municipal, etc.)
- Calendario escolar
- Razón profesor, personal directivo y administrativo por alumno
- Número de estudiantes por curso y por escuela
- Infraestructura básica (bancos, aulas, biblioteca, entre otros)
- Recursos instruccionales (textos, apoyos materiales, entre otros)
- Incentivos para directivos y profesores

Alumnos:

- Género, edad, etnia, nivel socioeconómico.
- Historia escolar (rendimiento, variables afectivas).

Currículo planeado:

- Ciclos, asignaturas, contenidos, objetivos.

Profesor:

- Género, edad, etnia.
- Formación previa, posgrado, perfeccionamiento en servicio.
- Años de experiencia, remuneración, jornada de contratación.
- Compromiso y satisfacción personal.

Indicadores de procesos

Política social-educativa (gestión institucional)

- Sistema de administración (centralizado-descentralizado)
- Sistema de asignación de recursos
- Incentivos a la innovación en las escuelas; supervisión y apoyo técnico

Familia:

Compromiso educacional familiar. Participación de los padres y satisfacción
 Guía, apoyo y estímulo académico a los hijos
 Expectativas y aspiraciones de padres respecto a educación de los hijos
 Hábitos, rutinas y participación de niños en tareas del hogar

Escuela (gestión escolar):

Clima organizacional
 Liderazgo académico, funciones, expectativas y atribuciones del director
 Rotación de personal docente y ausentismo laboral
 Niveles de participación de la comunidad educativa

Alumno (hábitos y destrezas):

Asistencia a clases. Puntualidad. Tiempo en la tarea. Participación en clases

Currículo realizado:

Días y horas efectivos de clases. Cumplimiento de metas. Dificultades en el desarrollo del currículo.

Profesor (gestión pedagógica):

Estilo de enseñanza. Empleo de medios y materiales de apoyo. Asignación y revisión de tareas. Oportunidad de aprendizaje
 Expectativas y atribuciones de los profesores
 Frecuencia de evaluaciones. Disciplina escolar
 Distribución del tiempo de profesores

Indicadores de productos (niveles y calidad de aprendizaje)

Logro de contenidos mínimos en Matemáticas, Lengua
 Nivel de destrezas cognitivas en cálculo, lengua
 Nivel de autoestima académica
 Años para completar el primer ciclo básico
 Tasas de deserción, repetición

Es obvio que este listado es sólo un esbozo y no un sistema de indicadores plenamente desarrollado. Se mezclan indicadores precisos, los cuales coinciden con los de otros sistemas, con ciertos datos que pueden servir para construir indicadores pero que, en sí mismos, evidentemente no lo son, como, por ejemplo, las disposiciones sobre la escolaridad obligatoria

y otros puntos del rubro de política social-educativa.

A partir del año 2000, a resultas de una decisión de la segunda Cumbre de las Américas, que tuvo lugar en Chile en 1998, un grupo del Ministerio de Educación del país, con apoyo de la OREALC, propuso un Proyecto Regional de Indicadores Educativos para la región.

El proyecto involucró a la mayoría de los países, y se basó en un marco conceptual que tomó en cuenta las aportaciones de los proyectos INES y WEI de la OCDE (véase supra), pero también las circunstancias de la región. Con un marco sistémico, el marco del proyecto latinoamericano en su primera versión, de enero de 2000, manejó cinco categorías, siete subcategorías y 28 indicadores, como sigue:

- *Contexto socioeconómico y descripción del sistema*
 1. Contexto demográfico, social y económico (tres indicadores)
 2. Descripción del sistema educativo (tres indicadores)
- *Recursos en educación*
 3. Recursos humanos (cinco indicadores)
 4. Recursos financieros (ocho indicadores)
 5. Recursos materiales
- *Desempeño del sistema. alcance y eficiencia de la educación*
 6. Alcance de la educación (cuatro indicadores)
 7. Eficiencia interna (tres indicadores)
- *Calidad de la educación.*
- *Impacto social de la educación (2 indicadores)*

1.7. Propuesta de indicadores educativos para el estado de Aguascalientes

Un subsistema de indicadores no es un simple listado de puntos yuxtapuestos. Los indicadores deben integrarse lógicamente, según un concepto de calidad preciso. Por una parte, están los *objetos empíricos*: personas (aspirantes, alumnos, egresados, desertores, maestros, directores) que se sitúan en *instituciones* (plantales, agrupados en zonas, ubicadas en municipios). Esos objetos empíricos no pueden aprehenderse como tales en su totalidad: de su inagotable realidad se deben precisar aspectos pertinentes para propósitos y obtener información sobre ellos. Debemos definir unos *objetos contruidos* los cuales permitirán obtener datos (pese a su etimología, no son simplemente datos, sino que



debieron construir) sobre ciertas dimensiones del objeto general de estudio, la calidad educativa: información sobre el contexto del sistema educativo, sus insumos y productos y ciertos procesos que tienen lugar en su seno. Con estos elementos se podrán construir relaciones con las cuales se llegará, finalmente, a las dimensiones del concepto de calidad: relevancia, eficacia, impacto, equidad y eficiencia. La información necesaria para construir los indicadores podrá provenir de varias fuentes:

- Datos estadísticos escolares derivados de procesos administrativos existentes (v.gr. formas 911) o por desarrollar (v.gr. estadísticas basadas en datos individuales) (DEE).
- Datos estadísticos derivados de otras fuentes, como el INEGI u otras dependencias (DEO).
- Resultados de aplicación de pruebas por parte de una instancia especializada (PRU).
- Resultados de otros instrumentos como cuestionarios para obtener datos sobre instalaciones y equipamiento de cada escuela (censo de planteles), encuestas de opinión entre padres de familia y población en general, etcétera (OTR).
- Resultados, debidamente integrados, de los procesos de evaluación de centros escolares que puedan realizarse por parte de los equipos de supervisión de Zona o por otras instancias, como la Fundación para la Excelencia Educativa (ESC).

Se sugiere incluir los indicadores siguientes, señalando la fuente de obtención de cada uno (DEE, DEO, etcétera), y destacando los que pueden obtenerse a corto plazo:

Indicadores de contexto

Aspectos demográficos y económicos

- | | |
|------|--|
| CO-1 | Porcentajes de personas de 5 a 29 años de edad en la población DEO. |
| CO-2 | Nivel socioeconómico promedio de la población DEO. |
| CO-3 | Porcentaje de población rural DEO. |
| CO-4 | Porcentaje de población con ingresos menores al salario mínimo DEO. |
| CO-5 | Índice compuesto de marginación DEO. |
| CO-6 | Problemática en entorno escolar (pobreza extrema, aislamiento geográfico, conflictividad, población migrante ESC). |
| CO-7 | Existencia de servicios de apoyo a escuelas |

por instituciones públicas o privadas ESC. C1 a C5 se obtendrán para la entidad y se desagregarán por municipio, localidad y, de ser posible, por AGEB. C6 y C7 se obtendrán por escuela, a partir de evaluaciones de centro.

Indicadores de insumos

Instalaciones y equipamiento:

- | | |
|------|---|
| IN-1 | Números de escuelas de cada nivel y modalidad (DEE). |
| IN-2 | Número de escuelas de organización completa, unidocentes (DEE). |
| IN-3 | Instalaciones y equipamiento de escuelas. Porcentaje bajo estándares mínimos (OTR). |
| IN-4 | Estado de conservación de instalaciones y equipo (OTR). |
| IN-5 | Disponibilidad de libros de texto (OTR). |
| IN-6 | Número de volúmenes en la biblioteca escolar (OTR). |

Maestros:

- | | |
|------|---|
| MA-1 | Número de maestros de preescolar, primaria, secundaria (DEE). |
| MA-2 | Nº de horas contratadas en secundaria (DEE). |
| MA-3 | Distribución de maestros por edad, sexo, estado civil (DEE). |
| MA-4 | Porcentaje maestros con diverso nivel de formación inicial (sin formación, normal sin licenciatura, con licenciatura, más) (ESC-OTR). |
| MA-5 | Porcentaje de maestros con x nivel de carrera magisterial (ESC-OTR). |
| MA-6 | Solicitudes de cambio de maestros en el año (ESC). |
| MA-7 | Salario inicial y final de maestros en relación con mínimos de profesiones (DEO). |
| MA-8 | Proporción de maestros de apoyo-director en relación con maestros con grupo (ESC-OTR). |

Personal de apoyo:

- | | |
|------|---|
| PA-1 | Personal de apoyo en escuelas preescolar, 1ª, 2ª (DEE-OTR). |
| PA-2 | Proporción de personal de apoyo en relación con maestros (DEE-OTR). |
| PA-3 | Porcentaje de personal activo en sistema |



Cualitativos:

- IM-2 Población con x nivel en pruebas (alfabetización funcional, aritmética, científica, computacional...) PRU
- IM-3 Empleo y desempleo por género-edad-nivel de escolaridad. DEO
- IM-4 Ingresos por género-edad-nivel de escolaridad. DEO
- IM-5 Prácticas culturales por género-edad-nivel de escolaridad. OTR
- IM-6 Participación en actividades cívicas por género, edad y nivel de escolaridad. OTR

Indicadores de procesos o cualitativos

- PR-1 Escuelas con evidencia de cumplimiento de normalidad mínima: asistencia de maestros y alumnos, horas efectivas de enseñanza (ESC)
- PR-2 Días perdidos al año por maestro, por diversos motivos (ESC)
- PR-3 Escuelas con evidencia de ambiente ordenado y productivo, sin problemas fuertes de violencia, conflictividad (ESC)
- PR-4 Escuelas con evidencia de procesos serios de planeación y evaluación que muestren liderazgo efectivo (ESC)
- PR-5 Maestros con x nivel de calidad, según evaluación de desempeño docente basadas en observación en aula (ESC-OTR)
- PR-6 Escuelas con actividades serias de capacitación de maestros (ESC)
- PR-7 Visitas del director o del supervisor al año (ESC)
- PR-8 Escuelas con otras actividades especiales, con la comunidad, de biblioteca (ESC)
- PR-9 Escuelas con índice compuesto de calidad (ESC)
- PR-10 Escuelas con reconocimiento de la Fundación para la Excelencia Educativa (OTR)

Varios de los indicadores de maestros, que anteriormente se consideraron de manera general, pueden manejarse en forma más fina en el nivel de cada plantel, por ejemplo:

- M-4 Porcentaje de maestros con x nivel de formación inicial -sin formación, normal, entre otros (ESC-OTR)

- M-5 Porcentaje de maestros con x nivel de carrera magisterial (ESC-OTR)
- M-6 Rotación anual de maestros. Solicitudes de cambio anuales (ESC)
- M-8 Proporción de maestros de apoyo (incluyendo al director) en relación con los maestros con grupo (ESC-OTR)

Indicadores de equidad

Los indicadores de insumos pueden dar lugar indicadores de equidad si se desagregan a nivel municipal, de localidad o de escuela; asimismo, se analizan las diferencias en cuanto a dotación de algunos recursos. Otros pueden ser ciertos indicadores de cobertura y del esfuerzo educativo relativo:

- EQ-1 Población de cinco años que no asiste a preescolar (DEE-DEO)
- EQ-2 Población de 6 a 14 años que no asiste a la escuela (DEE-DEO)
- EQ-3 Población de 6 a 17 años que no asiste a la escuela (DEE-DEO)
- EQ-4 Alumnos por grupo en 1° de primaria y resto de grados (DEE)
- EQ-5 Alumnos por docente con grupo (DEE)
- EQ-6 Alumnos por maestros con y sin grupo (DEE)
- EQ-7 Alumnos con beca u apoyo similar (OTR-ESC)
- EQ-8 Alumnos atendidos en educación especial (DEE)

Indicadores de gasto educativo y eficiencia

- GE-1 Gasto educativo total, absoluto y como porcentaje del PIB. DEO)
- GE-2 Gasto educativo público total y porcentaje del gasto público total (DEO)
- GE-3 Gasto educativo privado total y por fuente (DEO)
- GE-4 Gasto por alumno absoluto y como porcentaje del PIB *per capita* (DEE-DEO)
- GE-5 Gasto educativo total por objeto del gasto: sueldos en general y de docentes, gasto corriente, inversiones (DEE-OTR)
- GE-6 Costo por egresado de cada nivel (DEE-OTR)

No todos los indicadores se pueden obtener desde los primeros momentos de la operación del Sistema.

Para optimizar el esfuerzo necesario y obtenerlos es conveniente centrar la atención en aquellos que satisfagan mejor los siguientes criterios: *consistencia técnica* (y así la información ofrecida sea confiable); *disponibilidad* (priorizando aquellos que resulten más económico manejar, porque se puedan obtener de fuentes preexistentes o mediante esfuerzos especiales menos pesados); *utilidad*, seleccionando los referentes a variables importantes y, de preferencia, de las llamadas alterables, puesto que la información proporcionada puede llevar más fácilmente a acciones de mejoramiento); e *interés*, en el sentido de dar prioridad a aquellos indicadores que cuenten con más probabilidades de llamar la atención de los actores, las autoridades y la sociedad sobre determinados aspectos del sistema educativo que se consideren de importancia.

Teniendo en cuenta el criterio de disponibilidad, se considera que el SEE podría comenzar manejando los indicadores resaltados en negritas en el listado anterior, como se indicó ya. En cuanto a los del rubro de eficacia, su disponibilidad es inmediata en la forma de cohortes aparentes; en relación con cohortes reales, en cambio, sólo podrán tenerse cuando opere el sistema de clave única. En la lista de indicadores pueden distinguirse algunos que contienen información más particular y otros, de carácter más integrador o sintético, que resumen varios del tipo anterior y ofrecen información más agregada. En la formulación precisa de cada indicador deberá cuidarse la univocidad de la terminología y la claridad de la definición para evitar ambigüedades.

Dado que el sistema de indicadores se concibe como permanente, deberá dar origen a series temporales consistentes. Si se modifica la definición de un indicador en algún momento las series que lo incluyan deberán corregirse en consecuencia. El sistema deberá incluir la definición expresa de puntos de referencia normativos, parámetros o criterios con los cuales se comparen las cifras obtenidas empíricamente para llegar a juicios de valor; esos criterios podrán ser medias estatales o nacionales, estándares propios, nacionales o internacionales, la situación del sistema en el pasado, o las metas establecidas para el futuro.

Teniendo en cuenta la dimensión de equidad del concepto de calidad, deberá atenderse con cuidado el establecimiento de criterios diferenciados, cuando sea procedente, para que las decisiones puedan

tomarse sin perjudicar a las personas o instituciones menos favorecidas. Un caso de especial importancia es el relativo al punto de referencia para hacer juicios de valor sobre los niveles de aprendizaje de una determinada escuela.

Suele proponerse el criterio del *valor agregado*, con mediciones del nivel de aprendizaje al principio y al fin de un grado, ciclo o nivel, para apreciar los avances logrados por un alumno, grupo o institución. Sin embargo, y además de la gran dificultad técnica de este tipo de mediciones, no se resuelve la cuestión de la equidad, puesto que la desventaja relativa de ciertos alumnos no sólo se refiere al volumen de conocimientos, sino también, y tal vez más, a la capacidad de aprendizaje.

Por ello se recomendarán acercamientos como los de Francia, en donde se comparan los niveles alcanzados por una escuela con los deseados en función de las características de su alumnado. Similar y más sencillo es el acercamiento del estado de California, donde cada escuela es comparada con otras de características similares en cuanto a alumnado y entorno (Véase De Landsheere, 1994: 24; Kaagan y Cooley, 1989).

También es interesante el enfoque utilizado en el estudio de Zimbabwe, en donde se analizó la varianza intraescuela y la interescuelas (véase Ross, 1992). Puede usarse también el índice de productividad desarrollado por la IEA, en el que se relaciona el nivel de resultados obtenidos en pruebas por ciertos alumnos con la proporción del grupo de edad escolarizado (*productivity yield*, véase De Landsheere, 1994: 105-106). Cuando sea pertinente, los datos se obtendrán y manejarán teniendo como unidad de análisis al individuo (alumno, maestro). En los demás casos la unidad de análisis será la escuela singular, que constituye el nivel clave de implementación de las políticas y acciones educativas, y debe serlo también para su evaluación. También en el caso de la información que se obtenga sobre individuos, para efectos del sistema de monitoreo, deberá agregarse para cada establecimiento. Las decisiones sobre personas nunca deberán tomarse únicamente con base en las informaciones de este sistema; si bien, será lógico que quienes deban tomarlas podrán utilizar estos datos como parte de los elementos que deberán tomar en cuenta. Los indicadores sobre maestros, éstos aparecen en el



apartado de insumos, podrían constituir un subsistema independiente, al igual que los dos los cuales se refieren al resto del personal del sistema educativo.

La noción de calidad se entenderá, en este caso, como el logro creciente de profesionalismo por parte de cada tipo de actores, medido a través de indicadores específicos para cada uno. El subsistema de evaluación de personal educativo debe monitorear el proceso de profesionalización en los diversos niveles del sistema. Se sugiere desarrollar un sistema que mejore lo que actualmente se hace en el marco del sistema de Carrera Magisterial, siguiendo los lineamientos de la literatura y, más adelante, las investigaciones nacionales sobre enseñanza efectiva, además de ampliar dicha evaluación para los niveles de directores de escuela, supervisores y funcionarios. Será importante incluir al personal de las escuelas normales.

En lo relativo a escuelas y a conjuntos de escuelas, como zonas u otras, la información que ofrezca el sistema será crucial para dicha toma de decisiones. Sin embargo, es claro que no es el personal técnico que opere el sistema de

monitoreo el que deberá tomarlas, sino las instancias jerárquicas facultadas para ello, del director de cada escuela a la máxima autoridad estatal, pasando por los niveles intermedios. Estas instancias utilizarán, lógicamente, la información que proporcione el sistema de monitoreo, pero siempre deberán tomar en cuenta otros elementos. La prudencia, ese conocimiento que fundamenta las acciones, no puede ser substituida con ningún sistema de indicadores. Por otra parte, será necesario que la información producida por el sistema se difunda en forma amplia y diversificada, teniendo en cuenta la gama de posibles usuarios: de manera sintética, clara y comprensible para el público en general; de manera suficiente para los diversos niveles de tomadores de decisiones; y con la posibilidad de difusión utilizable por investigadores o especialistas en evaluación. En todos los casos deberán cuidarse las normas legales y éticas relativas a la protección de la privacidad de las personas. Las observaciones que se acaban de hacer con respecto al Subsistema de Indicadores de Calidad del SEE se aplican también, con las debidas salvedades, a los otros dos subsistemas.



2. El Subsistema de pruebas de aprendizaje

El sistema de indicadores de la OCDE incluye un apartado de resultados educativos, con un rubro dedicado a resultados individuales otro a resultados colectivos del sistema y un tercero a los resultados en el mercado laboral. Esto es lógico, ya que una definición mínimamente exigente de calidad educativa no puede limitarse al estudio de los insumos de un sistema, sino que debe incluir el análisis de sus productos. Se sabe también que la información sobre insumos es la más fácilmente disponible, ya que puede encontrarse en archivos administrativos en todas las instituciones. En cambio, éstos rara vez podrá encontrarse información previamente existente sobre resultados que sea válida y confiable. Por ello, si se quiere tener indicadores sintéticos de resultados es necesario desarrollar mecanismos de elaboración y aplicación masiva de instrumentos de evaluación confiables, a no ser que se tenga la suerte de poder aprovechar los resultados de trabajos de esa naturaleza desarrollados en forma independiente, como fue el caso de la OCDE, que pudo aprovechar los trabajos de la IEA.

Los maestros evalúan regularmente a sus alumnos en varias formas, llegando finalmente a la asignación de una calificación *sumativa* (letra o nota numérica) éste significa la aprobación o reprobación de la materia o el grado y, en general, condiciona el paso al nivel o grado siguiente. En general estos procesos de evaluación los realiza cada maestro de manera independiente y sin utilizar instrumentos o procedimientos estandarizados o unificados. Esto es explicable y tiene ventajas y permite que la evaluación tenga en cuenta un gran número de apreciaciones finas casi imperceptibles que el maestro tiene del alumno a lo largo del curso, y posibilita la adecuación de la evaluación a las condiciones únicas e irrepetibles de cada escuela, cada grupo, cada materia, cada maestro y cada alumno.

Sin embargo, estas ventajas tienen una contraparte correlativa, muy importante también: la imposibilidad de comparar de manera confiable los resultados de unos alumnos con otros, cosa indispensable en una evaluación de conjunto. Nos encontramos, pues, ante una antinomia difícil de resolver, la cual no es sino una variante más del clásico problema de la singularidad y la generalizabilidad que la investigación no puede eludir, y que en el caso de las pruebas de conocimientos o similares se concreta en los clásicos problemas de confiabilidad y validez, a los cuales se hizo referencia.

Todo instrumento para evaluar aprendizajes enfrenta la dificultad de compaginar profundidad y finura con aplicabilidad en gran escala y comparabilidad. Instrumentos que no se elaboren con las reglas del arte no son estrictamente comparables. Si reactivos y pruebas no están bien calibrados con técnicas estadísticas rigurosas; puntuaciones iguales pueden corresponder a niveles de aprendizaje distintos. Aplicar pruebas idénticas para evitar problemas de comparabilidad sólo es factible en pequeña escala. Mediciones de muestras o poblaciones importantes, repetidas en el tiempo, requieren de versiones diferentes, pero estrictamente equivalentes de las pruebas, lo que sólo puede lograrse mediante el uso de las técnicas estadísticas apropiadas. Por otra parte, la calidad de un instrumento podría mejorar, en principio, incrementando el número y/o la calidad de sus reactivos o ítems: un mayor número de reactivos permitiría teóricamente una mayor confiabilidad, al incluir, por ejemplo, varios ítems sobre un mismo tema. Y obviamente mejores reactivos hacen una prueba mejor.

La validez también puede beneficiarse, en principio, de la mayor extensión de una prueba, ya que así sería posible atender mejor las diversas áreas del cur-



rrículo, incluyendo no sólo los temas más elementales, sino también los aspectos más finos de aplicación de conocimientos a la solución de problemas reales. Sin embargo, al mismo tiempo, la mayor extensión de los instrumentos complica significativamente su aplicación, al grado de que, a partir de cierto punto, esta última se volverá inviable.

Las tendencias metodológicas recientes de elaboración y aplicación de pruebas buscan precisamente resolver estos problemas y no pueden ignorarse. Resulta paradójico que las áreas dedicadas a estas tareas desconozcan, con frecuencia, no sólo los avances metodológicos más recientes, sino también planteamientos que tienen ya décadas de vigencia, como las pruebas construidas con referencia a criterio. Por otra parte, la aplicación de pruebas de conocimientos no es la única forma para enfrentar el problema de la comparabilidad de las evaluaciones del aprendizaje de los alumnos, hechas por los profesores en forma independiente. Hay otra manera menos conocida, que se designa con el término moderación, o suavización en el sentido preciso: que puede verse en esta cita:

Cuando las escuelas evalúan el desempeño de los alumnos en cierta materia... es necesario tener en cuenta el hecho de que escuelas y maestros pueden tener estándares distintos al calificar... los procedimientos que se emplean para ajustar las calificaciones para corregir esa falta de comparabilidad entre instituciones se designan con el término moderación (moderation). Los procedimientos de moderación que se usan en tales casos pueden implicar, bien sea el uso de ajustes estadísticos, bien sea el uso de juicios basados en muestras del trabajo de los alumnos. (Keeves, 1994:85)

Los ajustes estadísticos pueden ser sencillos o complejos: si se conocen las calificaciones de todos los alumnos de cada escuela o grupo, en lugar de usar cada calificación *bruta* es muy fácil transformarla a un puntaje estandarizado como un puntaje *z*, que informa no sobre un supuesto desempeño absoluto, sino sobre la posición relativa de un alumno en la distribución de su escuela o grupo: si se encuentra en la media, muy por encima o por debajo de ella etcétera. Un ajuste más complejo requiere la aplicación de una, o varias, pruebas estandarizadas, que sirven como punto de referencia para calibrar los puntajes brutos.

Keeves describe la forma cómo en Suecia se controla la calidad de las notas que los alumnos reciben al final de la escuela obligatoria, como ejemplo de procedimientos de moderación que no recurren al ajuste estadístico de los puntajes: en ciertos momentos de los últimos tres años de escolarización, se aplican en todas las escuelas del Suecia pruebas estandarizadas de conocimientos en seis materias básicas. Las pruebas se califican en una escala de cinco puntos. Luego se comparan las medias de las calificaciones asignadas por cada maestro a sus alumnos, con la media obtenida por éstos en la prueba estandarizada de la materia. Si las medias coinciden, o si la diferencia entre ambas no es mayor a 0.2, se aceptan las calificaciones finales asignadas por el maestro sin corrección alguna. Cuando la diferencia entre las medias es mayor, los maestros del último grado obligatorio de la escuela, como colectivo, corrigen las calificaciones de los alumnos de suerte que la media se sitúe dentro del rango mencionado (Keeves, 1994:88-89).

Keeves comenta que estos enfoques pueden ser adecuados en contextos en los que las diferencias entre escuelas no son muy grandes, pero no en el caso contrario, que suele darse en los países en vías de desarrollo (cita expresamente al Brasil). Los procedimientos complejos, tanto con ajustes estadísticos como con el uso de juicios colectivos de los equipos docentes, como el de Suecia, requieren un periodo muy grande y recursos humanos, así como una organización sumamente eficiente, puesto que siempre suponen la aplicación de uno o varios instrumentos estandarizados, por lo que no son fácilmente aplicables en países grandes y pobres (Keeves, 1994:45-46).

La segunda manera, más conocida, para obtener información confiable sobre los resultados del aprendizaje, perfectamente compatible con el uso de procedimientos de moderación, consiste en la aplicación masiva de pruebas de aprendizaje estandarizadas.

El resto de este apartado es una revisión de la literatura sobre los antecedentes y la situación actual de los procedimientos para elaborar instrumentos de evaluación de aprendizajes de tipo objetivo, que puedan aplicarse en gran escala, y cuyos resultados constituyan variables con características de medición adecuados para someterlos a análisis estadísticos complejos que impliquen supuestos exigentes.

2.1. Los antecedentes de la medición educativa

Las bases de la teoría de la medición fueron puestas desde el siglo XVIII por Laplace y Gauss, con el estudio de los errores de las observaciones empíricas y la concepción de la curva normal. La aplicación de estas ideas al campo educativo comienza en el siglo XIX. Desde sus inicios se distinguen algunos países: Alemania, cuyas universidades eran los centros de formación académica más importantes del mundo, desde Guillermo de Humboldt y aún antes; Inglaterra, patria de los antecesores de las ideas hereditaristas y de las técnicas estadísticas modernas; los Estados Unidos, cuya pujante sociedad y sus innovadores sistemas educativos, del nivel básico al universitario, se apropiaron más que ninguno otro las nuevas ideas; en menor medida Francia y las regiones francófonas de Suiza y Bélgica, cunas de innovadoras corrientes pedagógicas.

Luego las difíciles circunstancias europeas antes de la Primera Guerra Mundial hasta después de la segunda hicieron que el desarrollo de la psicometría moderna se volviera algo tan típicamente norteamericano. En 1931, al escuchar que los participantes en un Congreso Internacional se referían a la psicometría como estadounidense, E. L. Thorndike, se vio en la necesidad de protestar, diciendo que “por el bien de la ciencia y por nuestro bienestar, sería preferible que las pruebas estandarizadas no fueran denominadas *exámenes estadounidenses* (Joncich, 1968, citado por De Landsheere, 1996: 68). En las páginas siguientes se sintetiza el proceso anterior.

2.1.1 Las aportaciones alemanas del siglo XIX

Al tomar posesión de la cátedra que había sido de Kant en la Universidad de Königsberg, Herbart solicitó al rector el establecimiento de una escuela experimental para poner a prueba sus ideas pedagógicas, pero consideró necesario aclarar que consideraba *un sueño insensato* la existencia de una ciencia experimental del comportamiento (Travers, 1983, citado por De Landsheere, 1996: 34).

Sobre las huellas de físicos y químicos, que en los siglos XVII y XVIII hicieron nacer sus respectivas disciplinas a la par que reducían la imprecisión de nuestros conocimientos de los fenómenos naturales,

desarrollando procedimientos ingeniosos para medir realidades aparentemente tan inasibles como la presión atmosférica o el peso del aire, en el transcurso del siglo XIX, estudiosos como Fechner y Helmholtz comenzaron a estudiar, en forma empírica, fenómenos fisiológicos como la sensación y la percepción (sensopercepción), creando lo que se llamó la psicofísica (de Landsheere, 1996:37). En 1879, Wundt, alumno de Helmholtz, establece el primer laboratorio de psicología llamada experimental. A partir de allí, dado el número de estudiosos de diversos países que se formaron en ese laboratorio, se difunden a todo el mundo las nuevas ideas. (De Landsheere, 1996: 37)

2.1.2. Las aportaciones inglesas

En 1864 Fisher propuso escalas simples (*Scalebook*) para *evaluar las aptitudes, capacidades y conocimientos de los alumnos...*, e introduce la Estadística en la investigación educativa al incluir la noción de promedio de conocimientos de un grupo escolar (De Landsheere, 1996: 389)

En 1869 Francis Galton publica *Hereditary Genius* en la que usa la curva normal, cuya aplicación a fenómenos humanos había planteado Quetelet desde 1800, y maneja medias, medianas y varianzas; Galton introduce la idea de correlación e inventa nociones básicas en el campo de la medición psicológica y educativa como la de percentiles (De Landsheere, 1996: 80). Para entender la lentitud con que se extendió el uso de este tipo de enfoques en el país en donde surgieron, resulta interesante la observación que recuerda De Landsheere: en 1877 el Senado de la Universidad de Cambridge rechaza la solicitud de dos profesores, en el sentido de establecer un laboratorio de psicofísica, diciendo que “sería insultante para la religión pretender medir el alma del hombre”. El establecimiento de un laboratorio así deberá hacerlo Galton en forma privada en 1882; en 1892 un aventajado seguidor, Karl Pearson, a quien se debe el desarrollo del análisis de correlación y las bases del multivariado, publica su famosa obra *The Grammar of Science*; en 1901 Galton y Pearson fundan la revista *Biométrica* (De Landsheere, 1996: 392 y 394). Una generación más tarde, en 1925, Ronald Fisher, seguidor de los dos anteriores, publica su *Statistical Methods for Research Workers*, sistematizando el diseño experimental moderno (De Landsheere, 1996: 389-398).

2.1.3. Aportaciones francófonas

Se mencionó al belga Quetelet, y los franceses consideran que Bravais concibió el coeficiente de correlación medio siglo antes que Pearson. Aunque no puede contarse entre los iniciadores; el francés Alfred Binet realizó aportaciones de gran valor y clarificó ideas importantes en un momento temprano del desarrollo de la psicometría. Éstas hicieron que su famosa Escala Métrica de la Inteligencia, presentada en 1905, fuera seguramente el instrumento más influyente en su momento, no en Francia misma, pero sí en los Estados Unidos. Debe subrayarse que el mismo Binet, al publicar su obra *Las ideas modernas sobre los niños*, en 1909, escriba que su trabajo constituye un balance de treinta años de investigación experimental desarrollada en los Estados Unidos y Alemania (De Landsheere, 1996: 47).

2.1.4. Los desarrollos en los Estados Unidos

Por razones históricas, los Estados Unidos de Norteamérica elaboraron mucho antes que los países europeos, un sistema de educación de cobertura masiva, primero en los niveles básicos, desde principios del siglo XIX, con educadores como Horace Mann, y luego, en los niveles medio y superior, sobre todo desde la Ley Morrill de 1862. Por lo anterior, no debe sorprender que fuera en este país donde en forma más temprana se asiste al desarrollo de evaluaciones en gran escala, con la aplicación de pruebas impresas de aprendizaje de la historia a 500 escolares de 19 escuelas de Boston en 1845. Durante 1895 y 1897 J. M. Rice aplica una prueba de ortografía a 16 mil alumnos, y después una de aritmética a 13 mil y otra de comprensión de lectura a 8300. Estas encuestas se siguen aplicando posteriormente por el *National Council of Education* en 1908 y por la *National Education Association*, en 1911.

Por la misma época comenzaron a hacerse encuestas sobre deserción (E. L. Thorndike, 1907), estudios para determinar y comparar el gasto por alumno (Ayres, 1909); evaluaciones de escuelas experimentales en Indiana (Flexner, 1915-1918), etcétera. El aumento y la repetición de ese tipo de actividades llevó a la creación de oficinas dedicadas a este tipo de trabajos en los distritos escolares

más importantes, en las principales ciudades y universidades, lo que propició al surgimiento de la investigación educativa como actividad profesional, reflejada en la publicación del *Educational Research Bulletin* a partir de 1916 y del *Journal of Education Research* desde 1920, así como de la fundación, por los directores de un grupo de las oficinas de investigación existentes, en 1922, de la *Educational Research Association*. En 1930 se transformó en la *American Educational Research Association* (de Landsheere, 1996: 72-74 y 397-398).

Los estudiosos norteamericanos hicieron también aportaciones importantes, desde fines del siglo XIX, al desarrollo de la psicometría iniciado en Europa. En 1890 J. McKeen Cattell, discípulo de Wundt, con su artículo *Mental Tests and Measurements*, inventó la palabra *test* y publicó el texto fundacional de la psicometría. Más tarde (1916) W. A. Mc Call, alumno de Thorndike, se anticipa una década a R. Fisher en los planteamientos básicos del diseño experimental. En el campo de las pruebas de inteligencia, las desarrolladas por Binet son adaptadas por L. M. Terman en Stanford en 1916, y se extendieron con su aplicación por el ejército americano, con el *Army Test* de R. M. Yerkes, Terman y otros en 1917 (de Landsheere, 1996: 56-71).

2.2. Desarrollos teóricos y técnicos en el campo de las pruebas

2.2.1. La teoría clásica de la medición

Construyendo sobre las bases anteriores, a partir de los primeros años del siglo XX se desarrollaron los diversos aspectos metodológicos de lo que se conoce ahora como la teoría clásica de las pruebas (*classical tests theory*), a partir de la teoría de la confiabilidad y el modelo estadístico de las puntuaciones, con las nociones de puntaje verdadero, error de medida y confiabilidad de la prueba (*true score, measurement error & test reliability*, véase Keeves, 1997: 707); los trabajos pioneros fueron los del inglés Charles Spearman, discípulo de Galton, entre 1904 y 1913, y la obra del norteamericano Edward L. Thorndike *An introduction to theory of mental and social measurement*, publicada en Nueva York, también en 1904 (Martínez Arias, 1995: 40).

Martínez Arias se refiere a Du Bois para señalar que a partir de la cuarta década del siglo XX, con la aparición de publicaciones especializadas como *Psychometrika* (1935) y *Educational and Psychological Measurement* (1941), la teoría de los tests se separa de la evaluación y la psicología diferencial y que, en su versión clásica, puede considerarse completa con la aparición del libro clásico *Theory of mental tests* de Gulliksen, en 1950 (1995: 42). Es interesante señalar también que la primera edición del anuario *Mental Measurement Yearbook* del Instituto Buros es de 1938, y que la *Psychometric Society* se fundó a iniciativa de L. Thurstone en 1935. Además, fue durante 1946 cuando Stevens formuló la clasificación de los niveles de medición nominal, ordinal, de intervalo y razón (de Landsheere, 1996: 68).

2.2.2. Las nuevas teorías de la medición

Durante la segunda mitad del siglo XX el avance no se detuvo. En lo estrictamente psicométrico, los conceptos básicos mantienen su vigencia (confiabilidad, validez, etcétera), pero la manera relativamente ingenua en que los aplica la teoría clásica, es progresivamente enriquecida por los planteamientos más sofisticados de dos importantes teorías: la de la generalizabilidad y la de respuesta al ítem.

La teoría de la generalizabilidad, desarrollada por L. Cronbach y colaboradores, constituye una extensión de la teoría clásica, que atiende en forma más comprensiva y satisfactoria, la problemática de la confiabilidad, substituyéndola por la noción de generalizabilidad, que substituye el concepto de puntaje verdadero por el de puntaje del universo (*universe score*) y, en lugar de manejar el error de medición en forma global, identifica diversas fuentes posibles de error, o facetas, y logra detectar su influencia, gracias a técnicas estadísticas multivariadas como el MANOVA. Con antecedentes en las décadas anteriores, el trabajo fundacional fue publicado en 1963: *Theory of generalizability. A liberation of reliability theory* (Martínez Arias, 1995: 170). Una década después se publicó la obra que contiene la versión madura de la teoría (Cronbach, 1972).

La teoría de respuesta al ítem, o de las curvas características de un ítem, *intenta dar una fundamentación probabilística al problema de la medición de rasgos y constructos no observables (rasgos latentes), considerando al*

ítem, y no al puntaje global, como la unidad básica de análisis (Martínez Arias, 1995: 237-243).

Las primeras ideas de esta teoría, *ítem response theory*, se remontan a Binet y Simon (1905-1916); Thurstone (1925-1929); Lawley (1934); y Richardson (1936). En la década siguiente Ferguson se acercó al planteamiento de las curvas características de los ítems (1942), término acuñado por Tucker en 1945. Luego surgió la noción clave de rasgos latentes en varios ámbitos de las ciencias del hombre, comenzando con Lazarsfeld en sociología; su expresión en el campo de la psicometría no es otra cosa que la *teoría de respuesta a los ítems* (TRI), que en sus inicios se designaba más bien con la expresión de *teoría o modelos de rasgo latente*, y cuyo inicio formal se considera que fue el importante artículo publicado por Lord en 1952 con el título de *A theory of tests scores*.

Simplificando modelos estadísticos basados en la curva normal acumulada, desarrollados por Lawley, Tucker y Lord, en los años cuarenta, Birnbaum desarrolló en 1957-58 los modelos logísticos de dos y tres parámetros. En 1960 Rasch propuso el de un parámetro que, pese a su mayor simplicidad, o tal vez precisamente por ello, y con el entusiasta apoyo de Benjamín Wright (Universidad de Chicago), se extendió más rápidamente. La obra de Lord y Novick *Statistical theories of mental test scores* (1968) difunde una versión madura de la TRI.

Con la difusión a través de las computadoras, los complicados cálculos para la aplicación de la TRI, se volvieron accesibles, con lo que se ha impuesto como nuevo paradigma de la psicometría contemporánea, cuyas aplicaciones para el mejoramiento de las pruebas incluyen lo relativo a equiparación, comparaciones longitudinales, estudios de sesgo con análisis del *funcionamiento diferencial de los ítems*, desarrollo de pruebas adaptativas; mejores bases para el establecimiento de puntos de corte en pruebas con referencia a criterio, etcétera. Ya en 1985 el libro de Hambleton y Swaminathan: *Ítem response theory. Principles and applications*, mostraba la amplitud de las aplicaciones de la TRI (Cfr. Muñiz, 1997).

2.2.3. Otros desarrollos recientes

Las últimas décadas vieron otras novedades en el campo de las pruebas, en muchos casos apoyándose



en la TRI. Unas fueron sólo la precisión de nociones clásicas, como la de validez, que se concebía simplemente como “el grado en que un test mide lo que dice medir” (Garret, 1937), y luego se volvió cada vez más compleja, distinguiéndose validez de contenido, de criterio concurrente o predictivo y de constructo (Martínez Arias, 1995: 329-335) para concluir en una concepción unitaria, con varias fuentes de evidencia de la validez. Pero además hubo diversas innovaciones metodológicas y prácticas importantes:

- Técnicas para valorar el sesgo de un ítem o instrumento, en relación con variables como género o grupo étnico de los sustentantes: análisis de tablas de contingencia y de varianza, diagramas de dispersión de coeficientes delta, dificultades transformadas y procedimientos basados en la TRI, *differential item functioning*. (Martínez Arias, 1995: 577-612)
- Tests los cuales miden el grado en que un sustentante alcanza un nivel de *maestría* previamente definido, en lugar de determinar su posición en relación con los demás sujetos: tests referidos a criterio (*criterion referenced tests*), según la terminología introducida por R. Glaser en 1963, desarrollados luego por Popham, Husek y Hambleton. (Martínez Arias, 1995: 653-693)
- Desarrollo de *ítems* que no se limiten a presentar alternativas estructuradas de respuesta entre las cuales debe escoger el respondiente (opción múltiple, falso-verdadero), sino que requieran de la elaboración de la respuesta, acercándose a las tradicionales preguntas tipo ensayo, pero con modalidades que permitan una calificación objetiva, con procedimientos de jueceo.
- Pruebas de ejecución o desempeño, en donde las acciones requeridas para responder la prueba se aproximan lo más posible a las que deberá ejecutar una persona al desempeñar en la vida real una actividad determinada, a lo cual se refiere la expresión *authentic testing*.
- Pruebas adaptativas por computadora, que no son simplemente pruebas tradicionales en un soporte moderno, también permiten ajustar el conjunto de ítems al nivel de conocimientos de cada sustentante, presentándole sucesivamente preguntas de un grado de dificultad que depende

de sus respuestas anteriores, con lo cual se puede aumentar la eficiencia de la prueba.

- Aprovechamiento del análisis factorial para identificar factores subyacentes a varios ítems y facilitar la definición de constructos, y el uso intencional de ítems multidimensionales.
- Muestreo matricial (*matrix sampling*) éste permite aplicar pruebas que cubran mejor ciertos dominios, cuando no se desea obtener resultados confiables a nivel individual sino grupal.
- Las *adaptaciones*, o variaciones controladas de los procedimientos de aplicación de una prueba, para atender las condiciones particulares de ciertos sustentantes, como los afectados por determinadas discapacidades (*acomodations*).

Estos y otros avances permiten hablar de una nueva generación de pruebas, muy distintas de las de los años cincuenta, que eran típicamente preguntas de opción múltiple aplicadas en forma controlada, cuyos resultados se analizaban con la teoría clásica. Actualmente existe una rica gama de pruebas (adaptativas, de ejecución, libres de sesgo) las cuales atienden necesidades de sustentantes especiales, que se analizan con base en la TRI, la teoría de generalizabilidad y las aplicaciones psicométricas de los sofisticados avances de la estadística multivariada, como los modelos de ecuaciones estructurales, los de variables latentes y los diseñados para manejar variables categóricas.

2.3. Experiencias de evaluación en gran escala en otros países

La tradición de aplicar pruebas escritas a los alumnos, de un establecimiento, para evaluar diversos aspectos de su aprendizaje es antigua, aún sin tomar en cuenta los casos esporádicos o las experiencias no sistemáticas, como los exámenes de Sócrates o de los sofistas.

El sistema de exámenes imperiales de China fue establecido formalmente en el año 606 de nuestra era. En Europa se considera que los iniciadores de la tradición moderna fueron los jesuitas en sus escuelas, desde la última parte del siglo XVI. En el siglo XIX varios de los grandes sistemas europeos implantaron exámenes nacionales para el acceso a las universidades, que han llegado a ser parte inevitable de los sistemas escolares respectivos. Es el caso del *abitur* ale-

mán, del *baccalaureat* francés, del examen italiano de *maturità* o el de *matriculation* del Reino Unido.

Sin embargo, éstos exámenes, tipo ensayo, tienen en común el consistir en pocas preguntas (incluso una sola) y requieren respuestas amplias mediante las cuales es posible, en principio, evaluar no sólo los conocimientos particulares de los sustentantes en relación con ciertos temas sino, ante todo, su cultura general, su *madurez*, su capacidad para formular un discurso propio que integre elementos variados en forma coherente y original.

El número de estudiantes, en creciente aumento, en el contexto del dinámico sistema educativo norteamericano de fines del XIX, y principios del XX, puso en evidencia la dificultad de calificar de manera rápida y confiable grandes cantidades de esos exámenes *tipo ensayo*. Lógicamente fue en ese contexto donde se desarrolló con más vigor la alternativa que constituyen las pruebas llamadas objetivas, que suelen ser de respuesta abierta muy breve, o bien de respuesta preestructurada, con varias modalidades de opciones predefinidas entre las cuales el alumno debe únicamente escoger y marcar la correcta, a su consideración.

Por esas circunstancias, y otras a las que ya se hizo alusión, al presentar los desarrollos teóricos relativos a medición y psicometría, el desarrollo más importante de los acercamientos a la evaluación educativa mediante instrumentos objetivos de aplicación en gran escala, se ha dado sobre todo en los Estados Unidos. Por ello el primer inciso de este apartado se dedicó a revisar algunas de las experiencias más destacadas de dicho país, comenzando por las concernientes con la educación superior, con la experiencia del *College Board*, para después presentar lo relativo a la educación básica, con el ETS y el NAEP. En apartados posteriores se presentarán experiencias de otros países, y de alcance internacional.

2.3.1. Estados Unidos

Las pruebas del *College Board*

El nacimiento del *College Board*, originalmente denominado *College Entrance Examination Board*, el 17 de noviembre del año 1900, tuvo lugar en un momento en el cual el fuerte incremento de la proporción de

jóvenes que terminaban la educación media superior (*senior high school*), con la proliferación simultánea de instituciones de educación superior (*colleges*) de muy diverso perfil, que tenía lugar en esa época en los Estados Unidos, hacía muy complejos los procesos de selección de una institución y los trámites para ingresar a ella.

El primer historiador del *College Board*, Claude M. Fuess, señaló que la fundación de esta institución fue el primer intento organizado por *introducir la ley y el orden en una anarquía educativa que, a fines del siglo XIX, había llegado a ser exasperante, sin duda casi intolerable, para los directores de escuelas*, y añadió que había *un consenso preocupantemente reducido entre los colleges en cuanto al tipo de preparación en ciertas áreas de contenido y en cuanto a los estándares de desempeño que debían pedirse a los aspirantes* (Citado en Donlon, 1984: 1).

Si se tiene en cuenta que en 1900 aún no se desarrollaban los conceptos modernos de la teoría de la medición, ni las técnicas de la psicometría, ni en su versión clásica, no debe extrañar que las primeras pruebas del *College Board* fueran muy distintas de las actuales. Eran pruebas de tipo ensayo en nueve áreas cuya enumeración recuerda también las concepciones curriculares de una época que parece ya lejana: Inglés, Francés, Alemán, Latín, Griego, Historia, Matemáticas, Química, Física. Como las pruebas se aplicaban en distintos lugares, la comparabilidad se aseguraba haciendo en forma simultánea y asegurando la uniformidad de contenido, de condiciones de administración y de la calificación de las respuestas (Donlon, 1984: 1).

Las pruebas eran elaboradas por comités de maestros reconocidos. La primera aplicación tuvo lugar siete meses después de la fundación del *College*, el 17 de junio de 1901, e implicó a 973 aspirantes, en 69 lugares diferentes, con una fuerte concentración en la Universidad de Columbia. La mayoría de los aspirantes presentaron las pruebas de las nueve áreas mencionadas, para un total de 7 mil 889 pruebas, que fueron calificadas por comités de revisores reunidos en la biblioteca de la Universidad de Columbia. En 1902 se aplicaron pruebas a mil 362 aspirantes para ingresar a alguna institución de educación superior, y en 1910 el número llegó a 3 mil 731. Parece interesante mencionar que muchas instituciones, tanto *high schools* como *colleges*, veían las actividades del na-



ciente *College Board* como una amenaza para su autonomía; una consecuencia de eso fue el abandono de las etiquetas calificativas asociadas inicialmente a los puntajes numéricos, que eran excelente para puntajes de 90 a 100; *bueno* de 75 a 89; *dudoso* de 60 a 74; *pobre* de 40 a 59; y *muy pobre* menos de 40. (Donlon, 1984: 2)

Debido al desarrollo de las técnicas psicométricas en el campo de las pruebas de inteligencia, sobre todo para uso del ejército norteamericano durante la Primera Guerra Mundial, el *College Board* estuvo en condiciones para decidir, en 1925, el desarrollar pruebas del tipo que después se difundió como de aptitud, en contraposición a las de conocimientos, buscando ir más allá de la memorización de datos aislados, acercándose a la evaluación de habilidades intelectuales básicas de tipo general, sin pretender medir la inteligencia.

Encomendado su diseño a especialistas como Robert Yerkes y Henry Moore, quienes habían adquirido gran experiencia con el *Army Test*, y bajo la coordinación de Carl Brigham, la prueba que aún se llama *Scholastic Aptitude Test* (SAT); se gestó a partir de abril de 1925, y vio la primera luz el 23 de junio de 1926, cuando se administró a 8 mil 040 sustentantes. Las nueve subpruebas iniciales (definiciones, problemas aritméticos, clasificaciones, lenguaje artificial, antónimos, series, analogías, inferencias lógicas y lectura de párrafos), se redujeron a siete en 1928, y en 1929 a seis; que se agruparon en dos secciones: de aptitud verbal y numérica respectivamente (Donlon, 1984: 2).

Como no se habían desarrollado aún las técnicas de equiparación de versiones, la constatación de el porcentaje de sustentantes con puntajes considerados aprobatorios variaba de un año a otro en forma considerable, hizo pensar: lo que cambiaba en realidad era el grado de dificultad de la prueba, y no el nivel promedio de los sustentantes, que se podía suponer más estable. Por ello, a principios de los años treinta se decidió establecer una proporción fija de aprobados, ajustando las puntuaciones del conjunto de sustentantes de cada año en consecuencia (Donlon, 1984: 3).

Hasta 1936 se hizo una aplicación anual; en 1937 se hicieron por primera vez dos: la tradicional de junio y otra en abril. Mientras las pruebas de junio se-

guían siendo de tipo ensayo, aplicadas en varios días; las de abril, desde su primera aplicación, se componían íntegramente de preguntas de opción múltiple. Ambas se calificaban en una escala cuya media era de 500 y su desviación estándar era de 100. La equiparación de versiones y el cuidado de la estabilidad de la prueba a lo largo del tiempo, comenzó a hacerse en 1941, fecha importante en la historia del SAT, ya que es precisamente la población de 10 mil 654 sustentantes de la aplicación de abril de dicho año, la que constituye el punto de referencia en relación con el cual se estandarizan los puntajes de todos los sustentantes del SAT, desde entonces hasta la fecha. A partir de 1942 también las pruebas aplicadas en el mes de junio de cada año, consistieron exclusivamente en preguntas de opción múltiple. En los años cincuenta el *College Board* realizó tres aplicaciones del SAT al año; en 1977 realizó seis, y en los años noventa doce. El número total de sustentantes anuales llegó desde fines de los años sesenta al millón y medio de personas, cifra que se ha mantenido desde entonces (Donlon, 1984: 3-8).

Desde sus inicios el *College Board* estableció un Comité Revisor, formado por educadores reconocidos, para supervisar todos los aspectos del desarrollo de sus instrumentos. Con los avances de la psicometría y el desarrollo de nuevos tipos de instrumentos éstos fueron sistematizándose y ampliándose progresivamente las medidas de control de la calidad. Influyó también la aparición de disposiciones legales surgidas de la preocupación de asegurar la equidad de las decisiones basadas, de alguna manera, en los resultados de pruebas.

Al aparecer el primer SAT, en 1926, y aunque se trataba de un conjunto de exámenes tipo ensayo, el *College Board* difundió el Manual de la prueba, del que procede el párrafo citado en el epígrafe. Desde entonces se utilizaron procedimientos de estandarización de las puntuaciones obtenidas por los sustentantes de cada aplicación; y la equiparación de versiones y el cuidado de la estabilidad de la prueba comenzó en 1941. En 1971 se publicó la primera edición de la obra llamada *The College Board Technical Handbook for the Scholastic Aptitude Test and Achievement Tests* que, como señaló su principal autor, William H. Angoff, no era un documento como el manual de 1926 y otros similares posteriores, en los que se proporcionaba in-

formación a los usuarios sobre la manera de utilizar correctamente la prueba para propósitos de orientación o selección, sino que se trataba de una obra en la que se sintetizaban numerosos trabajos técnicos y resultados de investigaciones, ofreciendo al especialista *toda la información necesaria para una evaluación técnica exhaustiva de la prueba*.

Con similar propósito, la segunda edición de la obra comprende diez capítulos en los cuales, además de la visión histórica de la que se toma la información de este apartado, se presentan los métodos psicométricos utilizados por el *College Board* en general y en particular en el SAT, en la prueba de Inglés escrito, en las pruebas de desempeño y en el Cuestionario descriptivo que se aplica a los sustentantes para recabar información sobre variables demográficas, escolares y socioeconómicas las cuales pueden usarse en investigaciones que traten de entender los factores que puedan explicar los resultados obtenidos en las pruebas por los distintos grupos de sustentantes. Se discuten estudios dirigidos al análisis de la validez de constructo, de contenido y predictiva: se analizan estadísticas descriptivas de los sustentantes y se presentan trabajos sobre el efecto del entrenamiento para la prueba y el posible sesgo de la misma. (Donlon, 1984)

El Cuestionario descriptivo se introdujo en 1971, al igual que los comités académicos de asesoría por áreas de conocimiento, y un comité de supervisión general con educadores, tanto de escuelas de nivel medio superior como superior (*high schools* y *colleges*). Otros datos que muestran el avance en la calidad metodológica del SAT, son los siguientes: el tamaño de la muestra de sustentantes que se consideraba adecuada para probar un ítem a ser incluido en la prueba era de 370 en 1961, y en 1975 subió a dos mil; la proporción de irregularidades en la aplicación que dan lugar a algún tipo de procedimiento de revisión es menor al 0.1%; la de los errores del proceso de calificación mecánica de las hojas de respuesta es prácticamente de 0. Desde 1953 los métodos de calificación incluyen una corrección por adivinación, con base en los resultados de estudios publicados ese año, los que fueron confirmados en otro estudio de 1973. A partir de 1982 comenzó a utilizarse la Teoría de Respuesta al Ítem para la equiparación de las puntuaciones de la prueba.

Como se ha indicado, otros cambios que el *College Board* debió hacer fueron el resultado de decisiones legales. Fueron importantes en particular las disposiciones aprobadas en 1980, en el estado de Nueva York, que hicieron obligatorio publicar cada prueba después de su aplicación, entregar a cada sustentante tanto las respuestas correctas a las preguntas de la prueba, como sus propias respuestas a éstas, entregar a las autoridades educativas estatales copia de todos los estudios relacionados con el SAT, así como dar información sobre lo que mide la prueba, sus limitaciones y la forma de calificarla. La obligatoriedad de publicar las pruebas aplicadas se ha ido precisando posteriormente, teniendo en cuenta el número de aplicaciones, por ser muy distinta la situación de las pruebas que se aplican a decenas o centenas de miles de sustentantes, en cada ocasión y la de tests con pocos sustentantes. Se han establecido períodos de publicación obligatoria en ciclos que van de cinco a ocho años. Durante los años noventa se desarrolló una nueva versión de la prueba, el *SAT II*, esta entró en su fase operativa en la segunda mitad de la década, en la que se introdujeron novedades importantes con respecto a la versión anterior, en particular ítems de respuesta construida, tipo ensayo u otro.

El SAT no es, por supuesto, la única prueba del *College Board*. Exámenes como el de selección para estudios de posgrado (*Graduate Record Examination, GRE*) o el de inglés para hablantes de otros idiomas (*Test of English as a Foreign Language, TOEFL*), se aplican anualmente a poblaciones de sustentantes comparables a las del SAT, y su desarrollo conoció seguramente procesos análogos.

El ETS y el NAEP

Desde la década de 1920, en las universidades de Princeton, EUA fue donde los trabajos de construcción de pruebas estandarizadas se desarrolló con fuerza, a cuyo personal académico pertenecía Carl Brigham, el principal autor del SAT. En 1948, la oficina de Princeton que se dedicaba a la elaboración de tests se separó de la universidad, aunque manteniendo vínculos estrechos con ella, constituyéndose formalmente como una entidad privada diferente. Esta unidad es el *Educational Testing Service*; sin duda el centro especializado en producción de instrumentos psicométricos.



cos más fuerte del mundo (De Landsheere 1986:150). Varios de los desarrollos técnicos y metodológicos más importantes en el terreno de la psicometría han sido el fruto del trabajo de investigadores del ETS.

Por otra parte, la difusión de las pruebas en gran escala y su uso para la evaluación de sistemas educativos, pasó por una etapa importante durante los años sesenta y setenta, con los trabajos que engloba la expresión *minimum competency testing*, enfoque adoptado en esa época por la mayoría de los estados norteamericanos, como pieza integral de los sistemas de evaluación establecidos con obligatoriedad legal a partir de esa época.

Como consecuencia de la preocupación por el deterioro de la calidad de las escuelas, tras el impacto del lanzamiento del *Sputnik* por la Unión Soviética, en 1957, el Informe Coleman, en 1966, y la aparición de una marcada tendencia a la baja de los resultados promedio obtenidos año tras año por los aspirantes a ingresar a la educación superior en el SAT, las legislaturas de algunos estados como California, Florida y Oregon, establecieron la obligatoriedad para aplicar regularmente a los alumnos de diversos grados de la educación preuniversitaria pruebas de conocimientos construidas en relación con determinados estándares mínimos, definidos en los planes de estudio respectivos. Éstos fueron los programas de *minimum competency testing*. Los opositores de esta tendencia argumentaron que no era evidente que hubiera realmente un deterioro en los promedios del SAT o similares, dados los cambios importantes en el alumnado de *High School* y, sobre todo, en la composición de la población de aspirantes a ingresar a las instituciones de educación superior. Sin embargo, los motivos de preocupación resultaron suficientemente fuertes para que la tendencia se fortaleciera, y de hecho, prácticamente todos los estados americanos implantaron de una u otra forma esas pruebas.

Según Baker y Choppin (1990), en 1982 una gran mayoría de estados (42/50) había implantado, o estaba implantando, programas obligatorios de esa naturaleza. Los mismos autores mencionan que, al generalizarse, los programas de pruebas de competencia mínima, muchas veces han sido implementados de manera deficiente, por lo que su impacto se ha reducido, y las expectativas depositadas en ellos no se han cumplido. Las demandas judiciales que acusan a

las pruebas de ser discriminatorias, sesgadas y poco fiables, se multiplicaron (Baker y Choppin 1990). En Latapí (1991) puede encontrarse una introducción a esta polémica, y la traducción de varios textos sobre ella publicados en *Phi Delta Kappan* en 1978-1979 por Boes, Bracey, Brikell, Cawelti, Ebel, McClung, Mcklenburger y Piphoo. En ese contexto fue creado en el *Educational Testing Service*, el programa llamado *National Assessment of Educational Progress*, con el propósito de tener información sobre el nivel de la educación a escala nacional. (véase Walberg 1990). Este programa, el NAEP, ha llegado a ser el punto de referencia obligado para los sistemas nacionales de pruebas para el monitoreo de la calidad educativa.

Hasta principios de los sesenta, en los Estados Unidos la responsabilidad educativa recaía en cada uno de los integrantes del pacto federal, por lo que no había ni un currículo nacional tampoco exámenes nacionales. La tradición de investigación educativa hacía que en muchos lugares se aplicaran pruebas estandarizadas, pero éstas podían ser muy diferentes y, además, su naturaleza normativa permitía ubicar a cada individuo en comparación con los demás, pero no daba elementos que permitieran saber qué tanto dominaba los contenidos que el currículo establecía. Por consiguiente, se carecía de datos confiables sobre el nivel de aprendizaje alcanzado por las escuelas y no había bases sólidas para un esfuerzo de calidad. En 1963, el Comisionado de Educación solicitó a la Comisión Carnegie, la realización de un estudio sobre esta problemática. Una comisión presidida por Ralph Tyler (*Exploratory Committee on Assessing the Progress of Education*), llevó a cabo un trabajo que terminó en 1969, y fue asumido para ser continuado en forma permanente por la *Education Commission of the States*. A partir de 1983 la realización de los estudios del programa NAEP fue confiada al ETS, el cual creó en su seno el *Center for the Assessment of Educational Progress* (CAEP). Con esta medida la calidad técnica de los trabajos del NAEP se incrementó. (Walberg, 1990:3478 y De Landsheere, 1994:50-51)

El NAEP ha desarrollado pruebas técnicamente muy avanzadas, con escalas basadas en la Teoría de Respuesta al Ítem, diseño matricial, etcétera, para la evaluación de diversas áreas curriculares. Según la importancia que se atribuye a éstas, la frecuencia de las aplicaciones varía, como sigue:

- la Lectura y las Matemáticas se evalúan cada dos años
- la Expresión escrita y las Ciencias cada cuatro
- otras áreas, incluyendo Educación Cívica, Geografía, Educación musical y en otras artes, formación profesional, etcétera, en forma irregular.

Las pruebas se aplican a grandes muestras (con muestreo matricial, o en espiral, pues no se pretenden resultados para el nivel individual, sino para el subsistema) de niños de nueve, 13 y 17 años, edades que corresponden a etapas importantes del proceso escolar. Las escalas de las pruebas permiten ubicar en un nivel a cualquier persona, y por tal las pruebas se aplican también a muestras de adultos, lo que permite tener una idea de el nivel de dominio de la lectura por parte de la población adulta, por ejemplo.

Desde el principio los resultados permitían análisis y comparaciones entre cuatro regiones del país EUA (noreste, sureste, centro y oeste), así como entre tipos de comunidades, grupos de alumnos con padres con diferente nivel de escolaridad, raza y sexo. Venciendo una fuerte resistencia, desde 1990 los resultados permiten también comparaciones entre estados. Las escalas del NAEP para evaluar las áreas curriculares permiten clasificar a la población en cinco niveles: rudimentario, 150 puntos; básico, 200; intermedio, 250; apto, 300; avanzado, 350, como sigue:

Niveles de la escala de Lectura:

- 150: Capacidad de realizar tareas de lectura simples y aisladas.
- 200: Capacidad de comprender informaciones aisladas o secuenciales.
- 250: Capacidad de buscar información específica, conectarla entre sí y generalizar.
- 300: Capacidad de comprender, resumir y explicar informaciones más complejas.
- 350: Capacidad de captar el contenido de textos especializados y sintetizarlo.

Niveles de la escala de Matemáticas:

- 150: Capacidad de efectuar sumas y restas elementales en forma mecánica.
- 200: Inicio de la comprensión y de la habilidad de las cuatro operaciones.
- 250: Dominio de las cuatro operaciones. Inicio de capacidad de solución de problemas.

- 300: Dominio de procedimientos de complejidad media con geometría y razonamiento.
- 350: Resolución de problemas en varias etapas. Dominio del álgebra básica.

Niveles de la escala de Ciencias:

- 150: El alumno conoce hechos científicos cotidianos.
- 200: Comprende principios científicos simples.
- 250: Aplica informaciones científicas básicas.
- 300: Analiza procesos y datos científicos.
- 350: Integra informaciones científicas especializadas.

No es difícil notar la relación con la taxonomía de Bloom de estas formulaciones breves de los niveles básicos de las escalas del NAEP. El valor de estos trabajos radica, como es lógico, en que van mucho más allá de la simple definición de los niveles. Las aportaciones del NAEP incluyen:

- La construcción de múltiples versiones equivalentes escaladas con todo rigor, de pruebas que permiten ubicar a un alumno en cierto nivel.
- La elaboración de los extensos bancos de reactivos calibrados que lo anterior supone.
- La aplicación masiva de las pruebas así desarrolladas.
- Los análisis comparativos que se vuelven posibles.
- Una inteligente difusión de los resultados, que los hace accesibles al público en general, evitando el uso de tecnicismos y facilitando la toma de decisiones.
- El apoyo a la escuela al difundir gran parte de los reactivos que ya no van a utilizarse, con lo que los maestros pueden construir instrumentos propios de calidad aceptable.
- El apoyo a los Estados para el desarrollo de sistemas propios vinculados técnicamente al NAEP, lo que permite aplicaciones de pruebas a muestras estatales mayores de las que se requieren para los estudios nacionales, para estudios locales más completos.
- El apoyo que todo lo anterior constituye para la comunidad científica, la cual puede beneficiarse de estos trabajos para sus propias investigaciones.

Durante los años ochenta, el ETS comenzó a realizar también estudios a nivel internacional, para



comparar el nivel de la educación norteamericana con el de los sistemas de otros países. Estos trabajos constituyen el *International Assessment of Educational Progress*, IAEP. (Véase Livingstone 1990).

La situación actual en los Estados Unidos

Al comenzar el siglo XXI, muchos países han empezaron a desarrollar sistemas de evaluación educativa en gran escala similares al norteamericano, por lo que el sistema educativo de los Estados Unidos de Norteamérica, ya no es tan excepcional como hace no mucho tiempo, aunque sigue siendo punto de referencia obligado por su peso específico y por la antigüedad, amplitud, variedad y solidez, de sus sistemas de evaluación. Por ello conviene terminar este punto comentado algunas opiniones recientes, según las cuales el sistema americano estaría apartándose de su propia tradición, al haber comenzado a abandonar los acercamientos de tipo objetivo y en gran escala, a la evaluación educativa, en favor de otras formas de evaluar, más cualitativas, y que se suelen manejar en el nivel de plantel.

En realidad, la situación americana reciente, lejos de caracterizarse por el abandono de las pruebas en gran escala, se distingue por combinar los acercamientos más diversos, y se está planteando con vigor el problema de la comparabilidad e integración de los resultados de un número creciente de sistemas de evaluación, desarrollados por cada una de sus entidades federativas, con referentes curriculares y criterios heterogéneos.

Desde 1990 se ha tratado de enfrentar este problema mediante las pruebas a muestras nacionales y estatales de alumnos del NAEP. Si bien se trata de pruebas avanzadas, que utilizan muestreo matricial de los contenidos, y que se aplican a muestras bien diseñadas de alumnos, mediante procedimientos bien controlados, no dejan de tener limitaciones claras. En particular, no permiten dar resultados a nivel de plantel, ni siquiera de distrito, y sólo cubren ciertas materias y grados: Lectura de 4° en 1992, 94 y 98; Escritura en 8° en 98; Matemáticas en 4° en 92 y 96, y en 8° en 90, 92, y 96; Ciencia en 8° en 1996.

Ante la dificultad de conjuntar cobertura amplia de contenidos, calidad técnica y posibilidad de analizar resultados a nivel de distrito y plantel; combina-

ción que ni siquiera un programa tan fuerte como el NAEP logra satisfactoriamente, se buscan otras soluciones: la Prueba nacional voluntaria (*Voluntary National Test*, VNT), propuesta por el Presidente Clinton en su mensaje sobre el Estado de la Unión de 1997, o un gran sistema de pruebas adaptativas computarizadas, aplicado por internet, que propone la *Rand Corporation*. (Klein y Hamilton, 1999)

Al comenzar el año 2002, el presidente Georges W. Bush promovió la aprobación de una nueva legislación educativa en el nivel federal, ésta se designa con la expresión: *que ningún niño se quede atrás*. Esta legislación implica innovaciones significativas en las políticas educativas norteamericanas en general, y en particular, en lo que se refiere a la evaluación del aprendizaje alcanzado por los alumnos. Para entender adecuadamente las motivaciones y los alcances de estos cambios, conviene recordar algunas cosas.

Los resultados obtenidos por los alumnos norteamericanos en las pruebas de PISA, difundidos en diciembre de 2002, al igual que había ocurrido antes en el TIMSS y otras evaluaciones de la IEA, sitúan a los Estados Unidos a media tabla entre los países desarrollados, lejos de los primeros puestos que les gustaría ocupar y que ahora acaparan países como Corea del Sur, Singapur, Japón, Finlandia o Canadá.

Las metas educativas fijadas por los gobernadores norteamericanos al comenzar la década de los noventa incluían el que, para el año 2000, sus estudiantes serían los primeros del mundo en Matemáticas y Ciencias. Un simple vistazo a los resultados de PISA muestran que eso no sucedió, y un sencillo análisis denota que eso se debe al bajo rendimiento de una parte importante del alumnado. En efecto, cientos de miles de alumnos de ese país obtienen resultados excelentes, que no desmerecen frente a los de los que obtienen los mejores alumnos de los países que ocupan los primeros lugares del *ranking* educativo mundial. El problema radica en otros centenares de miles de estudiantes cuyos resultados son muy malos.

A nadie sorprenderá saber que, como ocurría y a mediados de los años sesenta, según mostró el Informe Coleman, los primeros de estos dos tipos de alumnos son mayoritariamente del grupo de blancos-anglosajones-protestantes, lo cual implica además que suelen pertenecer a familias acomodadas, y se concentran en estados del noreste del país, como

Massachusetts o Minnesota; en cambio, los alumnos de bajos resultados, son masivamente negros e hispanos, pobres y, en muchos casos, de una lengua materna distinta al inglés. Es claro que los resultados promedio de los estudiantes americanos no podrán compararse con los de los actuales países líderes en este campo, en tanto los resultados de esos alumnos no mejoren substancialmente.

La nueva Ley Bush reconoce lo anterior, y propone diversas medidas para lograrlo. Un lugar importante en las nuevas políticas, lo ocupa el reforzamiento de los mecanismos de evaluación de la calidad educativa.

Aunque la evaluación es una actividad que se ha desarrollado temprana y extensivamente en los Estados Unidos, ciertas características del sistema educativo, de ese país, han limitado sus alcances e impacto; se trata en particular de la inexistencia de un currículo y unos estándares nacionales. Por ello la Ley Bush incluye disposiciones precisas en ese sentido: “todos los estados americanos deberán contar con estándares educativos claros, así como con sistemas estatales de evaluación alineados con dichos estándares, que deberán incluir la aplicación anual de pruebas de Inglés, Matemáticas y Ciencias a todos los alumnos de los grados 4to a 8avo”. Además, la participación de los estados en las pruebas nacionales del NAEP, que hasta ahora ha sido opcional, será condición obligatoria para que un estado pueda tener acceso a los importantes fondos federales, para apoyar programas de mejora educativa que la nueva legislación incluye.”

La Ley Bush establece un margen de tres años para que sus disposiciones sean exigibles, lo que ocurrirá a partir del año 2005. La amplitud del plazo, en el país que tiene la más fuerte tradición psicométrica, muestra la complejidad de la tarea, cuya implementación constituirá sin duda un cambio importante en la educación norteamericana.

2.3.2. Holanda y el CITO

No es fácil disponer de las condiciones que se dieron en los Estados Unidos y permitieron el surgimiento de instituciones y programas como el ETS y el NAEP. Entre los pocos países que cuentan con algo comparable se encuentran Australia y los Países Bajos. En este apartado se resume la experiencia de este último

país, con el Instituto Nacional para la Medición Educativa (CITO por sus siglas en holandés).

Creado por el Ministerio de Educación y Ciencia, en 1968, como una fundación no lucrativa, el CITO ha realizado desde entonces un trabajo muy variado y de calidad, y ha conformado una infraestructura humana y material de alto nivel. En 1987, en el marco de las preocupaciones recientes sobre la calidad de la educación, el CITO, recibió el estatus de institución de derecho público, con funciones ampliadas, gracias a la ley que define las estructuras de apoyo a la educación. Dichas estructuras comprenden tres elementos: por una parte, una inspección independiente, con un fuerte papel supervisor y asesor; en segundo lugar, una red de centros pedagógicos regionales y por subsistema, que ofrecen apoyo a las escuelas los maestros; y asimismo el propio CITO, como unidad de evaluación, investigación y desarrollo curricular.

La prueba de fin de la primaria (CITO Test)

Los trabajos más antiguos del CITO consisten en la elaboración y aplicación, a escala nacional, de una prueba para la evaluación individual de todos los alumnos al final de la escuela primaria. La educación básica neerlandesa comprende dos años de preescolar y seis de primaria, que también se designan como ocho años de primaria, de los 4 a los 12 años de edad, seguidos por cuatro a seis de secundaria, según las variantes de ese nivel; cuatro de esos años son obligatorios.

La prueba de fin de primaria es un test de aptitud (Lectura, razonamiento y Matemáticas), construido con referencia a la norma y de alta calidad técnica. Los resultados se dan a las escuelas, las que pueden usarlos para la admisión de los alumnos en los distintos tipos de educación secundaria y para su propia evaluación como instituciones, al recibir datos comparativos de las demás escuelas primarias con características similares. La prueba comprende 180 ítems y su aplicación tarda seis horas, en dos mañanas, tres horas cada una. Se aplica en las escuelas y los responsables de hacerlo son los maestros. No hay una supervisión especial. La prueba tiene un costo de aproximadamente 10 dólares por cada alumno. Cada año la prueba se aplica en un setenta por ciento del total de las escuelas, lo que representa aproximadamente



5 mil 400 planteles y 110 mil 000 alumnos, cada uno de los cuales llena dos hojas de respuestas. Los resultados se reportan tres semanas después, pues se requiere de una semana para recoger las hojas de respuestas, otra para su lectura por medios ópticos y su análisis, y una más, para imprimir los resultados.

Prepara la prueba un grupo de maestros, previa capacitación, trabajando en casa mediante pago adicional. Formulan unos 700 ítems, siguiendo un manual de especificaciones preciso. Los ítems se corrigen y mejoran en reuniones de formuladores; luego se pilotean, el año anterior a su uso final. Por último los técnicos del CITO seleccionan los ítems definitivos, a partir de los resultados del pilotaje y consultando a personas del inspectorado y de los centros de profesores.

El NAEP neerlandés

En la actualidad, los trabajos del CITO comprenden una gama mucho más variada de pruebas. Además de la de aptitud, hay un programa similar al NAEP de los Estados Unidos para evaluar la primaria, que se designa en inglés con las mismas iniciales, aunque con un significado ligeramente diferente: *National Assessment of Educational Performance* (en neerlandés las siglas son PPON). Como su equivalente americano, el NAEP-PPON está concebido para la evaluación del sistema nacional de educación y no para la evaluación individual.

Más amplio en su concepción que el americano, el NAEP-PPON cubre todas las áreas del currículo en un ciclo de cinco años. En 1992 se evaluaron Matemáticas y Música; luego se evaluó Lengua, en 1993 al final de la educación básica y en 1994 a la mitad de la misma, con Educación física; en 1995 Ciencias naturales y sociales; para 1996 se previó la evaluación del Inglés como lengua extranjera, educación para las Artes visuales y Educación vial. A diferencia, de la prueba de aptitud, las del NAEP-PPON son varias, éstas cubren todas las materias, incluyendo las que tradicionalmente se consideran más difíciles de evaluar, como la Educación artística. Para que sean relevantes y para que cubran los niveles taxonómicos superiores, incluyen aspectos no estructurados y de desempeño real.

Los instrumentos para evaluar la Educación artística, por ejemplo, incluyen un CD especialmente pro-

ducido para el propósito, con distintos fragmentos musicales sobre los que luego se pregunta al alumno, así como reproducciones de pinturas, etcétera. Esto hace que, inevitablemente, las pruebas se apliquen con más tiempo sean difíciles de corregir y calificar. Para efectos de investigación y evaluación del sistema, el CITO aplica los instrumentos del NAEP-PPON, en muestras de 200 a 300 escuelas, según la asignatura. Dado que se trata de aplicaciones más complejas que la prueba de aptitud: trabajar con muestras de tales dimensiones resulta manejable. Las escuelas se seleccionan aleatoriamente pero son libres de participar o no. Si una no desea hacerlo es reemplazada, también aleatoriamente. La aplicación y la corrección de las pruebas se hace por personal calificado.

El sistema de monitoreo de término de primaria

Una aplicación masiva de pruebas de esta naturaleza plantea problemas aparentemente insuperables. Sin embargo, el CITO ha desarrollado una estrategia que posibilita una aplicación en gran escala de pruebas complejas, conformando un sistema de monitoreo de la educación primaria, relacionado con el NAEP, pero de carácter formativo; esto es, para uso interno de cada escuela, con propósitos de mejoramiento.

En este caso, las pruebas son aplicadas y corregidas por los maestros de cada grupo: se hacen llegar a cada escuela los paquetes de las pruebas, que son aplicadas a todos los alumnos involucrados. Luego los profesores, de cada escuela, califican las pruebas de sus alumnos, para lo que han recibido entrenamiento. Utilizan materiales que facilitan dicha tarea y están incluidos en el paquete, como un programa de software especial, escalas para el análisis de los resultados, etcétera.

Un elemento esencial para que este enfoque pueda operar es que los resultados de las pruebas no deben enviarse al CITO, ni deben utilizarse para análisis, hechos centralmente, sobre la totalidad de los alumnos. Cada escuela es independiente para analizar los resultados de sus alumnos, y toma las decisiones que juzga adecuadas en consecuencia. Las escuelas tienen que reportar solamente los resultados de un alumno de cada grupo, seleccionado aleatoriamente con un procedimiento que define el

CITO. Estos alumnos forman una muestra nacional que puede usarse, comparando sus resultados con los del NAEP, como referencia para la evaluación de cada escuela, por las autoridades de cada una, y por los responsables del Sistema Educativo Nacional, para evaluaciones globales del sistema, sin referencia a personas o escuelas individuales.

Otras pruebas

Otros instrumentos desarrollados por el CITO, y que se utilizan en otros programas, como los de evaluación de la educación secundaria, incluyen pruebas diversas y especializadas para evaluar, por ejemplo, la aptitud física, el dominio oral del francés, el desempeño práctico en un laboratorio de química, etcétera. El programa más reciente del CITO y, probablemente el más impresionante, comenzó a ser puesto en operación en el año escolar 1992-1993, después de un período de preparación iniciado siete años antes. Se trata de un conjunto de materiales para todas y cada una de las materias que comprende el currículo de todos y cada uno de los grados de la educación básica. Los materiales para cada materia y cada grado incluyen

- Pruebas de excelente calidad para que el maestro las utilice con su grupo para diagnosticar el nivel de aprendizaje de sus alumnos de manera confiable, en relación con estándares nacionales bien definidos.
- Otros materiales que incluyen orientaciones sobre cómo tratar determinados problemas, así como material didáctico complementario con ejercicios, refuerzos, etc., que el maestro podrá utilizar con su grupo.

No se trata únicamente de pruebas, sino de paquetes que comprenden materiales para la evaluación de los aprendizajes y para el apoyo de los mismos, de conformidad con la misión del CITO, que comprende ambos aspectos. El uso de estos materiales es completamente voluntario, y su distribución es comercial: durante 92-93 el CITO lanzó al mercado los materiales, y cada escuela decidió si utilizará alguno(s) la adquisición de los materiales de una materia, como Matemáticas; o todos los de un grado, o cualquier combinación que se quiera. Cada escuela decide y compra los materiales que quiere.

El nivel profesional de los maestros, les permite aprovechar los materiales; el poder adquisitivo de las familias, permite que las escuelas puedan adquirirlos; la conciencia de la importancia de la educación que tienen maestros y padres de familia, y la participación efectiva de ambos, en la toma de decisiones en cada escuela; junto, por supuesto, con la calidad de los materiales, hicieron que, como esperaba el CITO, estén siendo utilizados por muchas escuelas y maestros. Se estima que durante el año escolar 94-95 la mitad de las escuelas de educación básica del país haya adquirido por lo menos, uno de los paquetes de materiales.

2.3.3. España y el Instituto Nacional de Calidad y Evaluación

En 1990, con la expedición de la Ley de Ordenación General del Sistema Educativo (LOGSE), España puso formalmente en marcha una importante reforma de su sistema educativo. La Ley subraya de manera destacada lo relativo a la calidad de la educación y, en ese contexto, asigna a la evaluación un papel importante. Para concretar esas ideas, un Decreto Real, del 18 de junio de 1993, creó el Instituto Nacional de Calidad y Evaluación (INCE) que comenzó a operar efectivamente en 1994. Su primer programa de trabajo contemplaba tres años (Plan de actuación 1994-1997), e incluía los siguientes renglones:

- Evaluación de la educación primaria.
- Evaluación piloto de la educación secundaria obligatoria.
- Evaluación piloto de la formación profesional.
- Sistema nacional de indicadores de calidad educativa.
- Estudios internacionales:
 - Tercer estudio internacional de Matemáticas y Ciencia (TIMSS).
 - Estudio internacional sobre lenguas en educación.
 - Estudio hispano-francés sobre enseñanza del Inglés.

El INCE concibe su misión relacionada con la evaluación del sistema educativo en su conjunto y no con la de alumnos, maestros o escuelas individuales. En lo relativo a la evaluación de escuelas (centros escolares), puede aportar elementos derivados de sus



trabajos a las escuelas y al inspectorado. Por otra parte, la naturaleza cuasi-federal de España hace necesario que el INCE tenga en cuenta para la realización de pruebas, y para su trabajo en general, a las autoridades educativas de las entidades con autonomía plena (como Cataluña, Galicia o el País Vasco), y a las regiones que dependen del Ministerio de Educación y Ciencia.

Ocho años después de su creación el INCE, con la experiencia de estudiosos que trabajaron previamente sobre temas de evaluación en el Ministerio de Educación, o en diversas universidades, ha realizado dos estudios de evaluación de la educación primaria, en 1995 y 1999, aplicando pruebas de Español, Matemáticas y conocimiento del medio a muestras de alumnos del último grado de la primaria (6°); Elaboró también dos estudios de la secundaria obligatoria, en 1996 y 2000. Realizó, además, estudios especiales, como el de la enseñanza de inglés; se encarga de la participación de España en estudios internacionales y maneja el sistema nacional de indicadores.

2.3.4. *El caso de Francia*

Aunque muy diferente, el caso francés va en la misma dirección. Tras la Revolución de 1789 y, sobre todo, desde las reformas napoleónicas de principios del siglo XIX, Francia se volvió el prototipo de una sociedad que se distingue por un estado fuerte y centralista en todos los ámbitos de la vida pública. La educación no es una excepción: es paradigmática la anécdota del ministro de educación que se jactaba de saber la lección que impartían en un momento dado, todos los maestros franceses sabiendo el día del año, mirando simplemente la hora en su reloj.

En contraste con lo que ocurre en países como los de América Latina que, a fines del siglo XIX y principios del XX, tomaron como modelo a la administración pública gala para el desarrollo de sus propios sistemas, en Francia el sistema funciona de manera razonablemente eficiente, gracias a un inspectorado respetado y a sistemas de evaluación de corte tradicional pero rigurosos, como el *baccalaureat*, todos los alumnos que terminan la enseñanza media superior deben aprobar para obtener el diploma respectivo, independientemente de las calificaciones que hayan obtenido en sus escuelas respectivas. Aplicado simul-

táneamente en toda Francia a todos los alumnos y en todas las materias del currículo, el famoso *bac* es aprobado solamente por un sesenta por ciento de los sustentantes en la primera oportunidad, y constituye la base para que la autoridad educativa nacional pueda tomar medidas las cuales aseguren un mínimo de calidad en todas las escuelas del país, en un enfoque que puede contraponerse al predominante muchos otros sistemas: el ideal Francés es que todas las escuelas sean buenas, no que haya unas excelentes, aunque otras tengan muchas deficiencias.

Si los valores subyacentes al sistema tradicional de evaluación de Francia, parecen interesantes, en el marco de una concepción educativa que destaca la importancia de la equidad, los enfoques técnicos parecen claramente superados, frente a los avances metodológicos y técnicos de la psicometría contemporánea. Es sintomático, en ese sentido, que en el seno del Ministerio de Educación de Francia la Dirección de Evaluación y Prospectiva esté desarrollando un sistema de evaluación de corte moderno, con pruebas avanzadas que se aplican a muestras nacionales de alumnos, que por el momento trabaja en forma paralela al sistema tradicional; éste es desarrollado por una instancia diferente (la *Maison des Examens*) y que posiblemente llegue a ocupar su lugar en el futuro.

2.3.5. *La evaluación de aprendizajes en América Latina*

Los países de América Latina, en general, se incorporaron a la tendencia mundial en lo relativo a evaluación educativa con cierto retraso. México constituye una excepción parcial en este ámbito: excepción, porque comenzó a aplicar pruebas desde principios de los años setenta; parcial, porque durante dos décadas las pruebas fueron de baja calidad y sus resultados no se difundieron, por lo cual en realidad no podía afirmarse que el país contaba con un verdadero sistema de evaluación.

Las otras excepciones regionales, también parciales, las constituyen Colombia y Costa Rica que, a través del Instituto Colombiano de Fomento de la Educación Superior (ICFES), y de la Universidad de Costa Rica, respectivamente, comenzaron a aplicar pruebas de aprendizaje a muestras de alumnos de educación básica desde los años setenta, pero igual-

mente sin que constituyeran un sistema nacional de evaluación en el sentido pleno de la expresión. En ese sentido la experiencia pionera fue la de Chile, cuyo gobierno encomendó a la Pontificia Universidad Católica que desarrollara un sistema de evaluación desde 1978. De manera más amplia es hasta fines de los años ochenta y, sobre todo, durante los noventa, cuando todos los países de la región, de manera acelerada, comenzaron a implementar sistemas de evaluación de concepción moderna y alcance nacional. En orden alfabético, la situación de los países de la región se presenta a continuación, a partir de Wolff, 1998; Tiana, 2000; e información directa de varios de los sistemas descritos.

Argentina

En 1993 el Ministerio de Educación del gobierno federal estableció un sistema nacional de evaluación, y desde esa fecha se aplican pruebas cuyo alcance y cobertura se han ido ampliando. Inicialmente se aplicaban a todos los alumnos del último grado de la educación primaria y la secundaria, e incluían solamente Matemáticas y Español; después se han extendido a otros grados y han incluido las Ciencias naturales y las sociales. Se aplican también cuestionarios a directores, maestros, alumnos y padres de familia, y se han hecho análisis sobre la influencia de los factores asociados con el rendimiento escolar, utilizando técnicas con estadísticas sofisticadas.

El Título IX de la Ley Federal de Educación de 1995 establece que el Ministerio de Cultura y Educación, así como las provincias y el municipio de Buenos Aires, deberán evaluar permanentemente el sistema educativo para garantizar la calidad de la educación. Se precisa que el Ministerio deberá informar anualmente sobre el particular a las comisiones de educación de las dos cámaras del Congreso de la Nación, y que *deberá convocar junto con el Consejo Federal de Cultura y Educación a especialistas de reconocida idoneidad e independencia de criterio para desarrollar las investigaciones pertinentes por medio de técnicas objetivas aceptadas y actualizadas* (Art. 48).

En el nivel federal las pruebas para evaluar la educación general básica (EGB), se aplicaron con regularidad por lo menos hasta 1999, cuando tuvo lugar el 7° Operativo Nacional de Evaluación Muestral de la EGB. En 1997 se aplicó por primera vez el Operativo Nacional

de Evaluación de Finalización del Nivel Secundario, repitiéndose por lo menos en 1998 y 1999. En los últimos meses del gobierno del Presidente de la Rúa se aprobó la creación de un órgano descentralizado, con cierta autonomía, para que se encargara de manera permanente de hacer las evaluaciones. La grave crisis por la cual atraviesa el país ha impedido que dicho organismo su funcionamiento efectivo, por lo que las actividades de evaluación siguen a cargo de dependencias del Ministerio de Educación y de las autoridades educativas de las provincias.

Bolivia

En este país se ha establecido un Sistema de Medición de la Calidad (Simecal) que comenzó en 1997 aplicando pruebas de rendimiento en los 3° y 6° de primaria, en las áreas de Lenguaje y Matemáticas, contemplando también el análisis de factores asociados. Con los resultados de las evaluaciones aplicadas en el año 2000 a los 1°, 3°, 6° y 8° de primaria, así como a 4° de secundaria, se estableció una Línea de Base como referencia fundamental para las evaluaciones posteriores. En 1999 se hizo también una Evaluación del Sistema de Formación Docente del país.

Brasil

Desde 1993, el gobierno del país más grande de América Latina, con más de 170 millones de habitantes, estableció un *Sistema de Avaliação da Educação Básica* (SAEB) que, sin embargo, no llegó a consolidarse en el periodo gubernamental cuando comenzó.

Desde los inicios del primer periodo de la administración del Presidente Cardoso el programa de evaluación educativa se reforzó, lo que llevó en 1997 a la creación de un órgano descentralizado, aunque no autónomo (una *autarquia*, en la terminología brasileña), para encargarse de la evaluación educativa: el *Instituto Nacional de Estudos e Pesquisas Educacionais* (INEP). A diferencia de los organismos de evaluación de la mayoría de los países, en las diversas regiones del mundo, el INEP comprende tanto lo relativo a educación básica y media, como lo que toca a la educación superior. Por ello, además de las pruebas para la educación básica del mencionado SAEB, el INEP aplica el *Exa-*



men Nacional do Ensino Médio (ENEM), incluso, un conjunto de exámenes de egreso para las principales carreras del primer nivel de la educación superior, excepcional en el panorama mundial de la evaluación educativa, que se denomina oficialmente *Exame Nacional de Cursos*, y se conoce popularmente como el PROVAO.

El SAEB es un programa que aplica cada dos años pruebas de Matemáticas y Lectura a muestras nacionales de alumnos de los 4° y 8° de la enseñanza primaria, y 2° y 3° de la secundaria. Además del año de 1993 se han aplicado tales pruebas en 1995, 1997, 1999 y 2001. Los resultados de la aplicación de 1999 se divulgaron en 2000, comprendiendo a más de 360 mil alumnos de más de 2 mil 100 municipios de los 26 estados y el Distrito Federal. El ENEM, por su parte, se aplicó a más de 390 mil egresados de la educación media superior, y es cada vez más utilizado como criterio para definir el ingreso a la universidad. En el año 2000 el PROVAO cubrió 18 carreras y fue aplicado a casi 198 mil egresados de instituciones de educación superior de todo el país, que ofrecieron en conjunto 2 mil 888 programas de esas carreras. Con un reconocimiento creciente por la elevada calidad técnica de su trabajo el INEP se encarga también del manejo del sistema nacional de indicadores educativos (véase INEP 2001).

Chile

Gracias al trabajo emprendido en 1978, Chile destaca por la introducción, desde 1981, de un sistema nacional de pruebas para la evaluación de la calidad de la educación básica, que en 1988 adquirió el nombre de Sistema de Medición de la Calidad de la Educación (SIMCE). Las pruebas del SIMCE se aplican a todos los alumnos de los 4° y 8° de la educación primaria en las áreas de Español y Matemáticas, y a una muestra de ellos en Ciencias Naturales, Historia y Geografía. El sistema chileno es reconocido por su calidad, y ha conocido mejoras técnicas importantes. Además del uso de los modelos avanzados de la TRI, se explora la introducción de preguntas de respuesta abierta o construida, además de las tradicionales de opción múltiple. El trabajo se desarrolla dentro del Ministerio de Educación, si bien involucra a especialistas de las principales universidades del país. Chile se distin-

gue también por ser el país latinoamericano que participa con mayor frecuencia en proyectos internacionales de evaluación.

Colombia

Algunos trabajos del ICFES en los setenta son antecedentes de la evaluación reciente, en particular las pruebas de ingreso a la educación superior. A partir de 1990, el ICFES comenzó a desarrollar pruebas para evaluar el desempeño de muestras nacionales de alumnos de los 3° y 5°, y después 7° y 9° de primaria, en Lengua y Matemáticas, y luego en ciencias naturales y sociales. El nivel técnico ha mejorado, introduciéndose la TRI. Se estudian también factores asociados.

Costa Rica

Con antecedentes en el trabajo del Instituto de Investigación para el Mejoramiento de la Educación Costarricense (IIMEC) de la Universidad de Costa Rica, el Ministerio de Educación Pública de ese país, con apoyo del Banco Mundial, emprendió un importante esfuerzo para consolidar la evaluación educativa, estableciendo un Centro Nacional de Evaluación para la Educación (CENE-EDU). Este centro, junto con el IIMEC, que sigue encargándose de la elaboración de la mayoría de las pruebas, aplica pruebas formativas en diversos grados y materias del currículum, así como pruebas sumativas en todas las materias a todos los alumnos que terminan el 9°, así como quienes terminan la enseñanza secundaria.

Ecuador

El Ministerio de Educación y Cultura estableció un Sistema Nacional de Medición de Logros Educativos, conocido como APRENDO, cuya primera aplicación de pruebas tuvo lugar en 1996. Las aplicaciones se hacen anualmente, a alumnos de 3°, 7° y 10° de educación básica, en Lengua y Matemáticas.

El Salvador

Como otros países centroamericanos, El Salvador participa en un proyecto promovido por la Organi-

zación de Estados Iberoamericanos (OEI), y la Coordinación Educativa y Cultural Centroamericana (CECC), para el Establecimiento de Estándares para la Educación Primaria en la región, que incluye un componente de evaluación.

Guatemala

Además de participar en el proyecto mencionado de la OEI y la CECC, Guatemala ha establecido un Programa Nacional de Evaluación del Rendimiento Escolar (Pronere); la primera aplicación de pruebas de este Programa tuvo lugar en 1999, cubriendo las áreas curriculares de Lectura y Matemáticas en los 3° y 6° de primaria.

Honduras

La Unidad de Medición de la Calidad de la Educación (UMCE) de la Universidad Pedagógica Nacional Francisco Morazán, por un acuerdo con la Secretaría de Educación, se encarga de las actividades de evaluación en este país. El primer levantamiento de datos tuvo lugar en 1997, cubriendo los 3° y 6° de primaria y las áreas de Español y Matemáticas. En 1998 se realizó el segundo levantamiento de datos, extendiéndose a los grados 2°, 3°, 4° y 6° de primaria, en las mismas áreas curriculares.

Nicaragua

Participa en el proyecto regional de la OEI y la CECC.

Panamá

Participa en el proyecto regional de la OEI y la CECC.

Paraguay

Con apoyo del BID, el Ministerio de Educación y Cultura ha establecido un Sistema Nacional de Evaluación del Proceso Educativo (SNEPE), cuya primera actividad se realizó en 1997 e implicó la aplicación de pruebas a los alumnos de 6° de primaria.

Perú

El Ministerio de Educación estableció la Unidad de Medición de la Calidad (UMC), que comenzó a aplicar pruebas de Lengua y Matemáticas a los alumnos de 4° de primaria en 1996; en 1998 las pruebas incluyeron también las áreas de Ciencias naturales y Ciencias sociales. La UMC cuenta con el apoyo de un grupo de investigadores independientes, el Grupo de Trabajo sobre Estándares y Evaluación (Grade).

Uruguay

En este país el órgano responsable del sistema educativo no es un ministerio; es una entidad que no depende del titular del ejecutivo, sino de un Consejo de personalidades designadas por el Congreso: la Administración Nacional de Educación Pública (ANEP). Esta entidad estableció en 1995 la Unidad de Medición de Resultados Educativos (UMRE), para encargarse de las actividades de evaluación, a partir de investigaciones realizadas en los años previos por el sociólogo Germán Rama, quien durante 1995 era el titular de la ANEP.

La UMRE, cuyo nivel técnico y eficiencia son reconocidos, realizó en 1998 una Evaluación Muestral de Aprendizajes, y en 1999 una Evaluación Censal de Aprendizajes que cubre a todos los alumnos de 6° de primaria, en las áreas de lengua y matemáticas. A partir de ese año las actividades de la UMRE, el proyecto de Medición de la Calidad de la Educación Primaria (Mecaep), cubre en forma censal anualmente 3°, 6° y 9° del ciclo básico de la educación.

2.3.6. Otras experiencias nacionales

No es raro escuchar opiniones en el sentido de que las pruebas objetivas están siendo abandonadas en los países avanzados, incluidos Estados Unidos, supuestamente por haberse extendido la conciencia de sus insuperables fallas radicales. Como se indicó, estas afirmaciones carecen de fundamento en lo referente al caso americano, en donde las pruebas están lejos de ser abandonadas. Lo verdadero, es que se vuelven cada vez más sofisticadas, para cuidar mejor la validez y la confiabilidad y evitar sesgos, los cuales pueden afectar los derechos de los sustentantes; también



es cierto que las pruebas se complementan cada vez con más frecuencia con acercamientos alternativos.

Curiosamente, los opositores norteamericanos a este tipo de instrumentos utilizan el mismo argumento, afirmando que las pruebas en gran escala están siendo abandonadas en otros países, y atribuyendo a la fuerte presencia de este tipo de evaluaciones, muchas fallas de las escuelas estadounidenses. Un artículo menciona cinco publicaciones americanas, aparecidas entre 1991 y 1996, en ese sentido (Phelps, 2000: 11).

Ese artículo señala que las afirmaciones referidas no presentan sustento sólido, sólo que se limitan a afirmar su posición señalando, por ejemplo, que Bélgica, Grecia, Portugal y España han eliminado ese tipo de exámenes nacionales. Cuando se sabe que, por el contrario, ha sido justo durante los años noventa cuando España comenzó a desarrollar ese tipo de evaluaciones, con la creación del Instituto Nacional para la Calidad de la Educación (INCE), hay razón para dudar de tales afirmaciones. El artículo de Richard Phelps analiza la situación de 31 países o provincias, con el resultado de que en la gran mayoría el uso de pruebas en gran escala, lejos de disminuir, está aumentando significativamente: 27 países o provincias han incrementado el número de pruebas estandarizadas en gran escala, en el cuarto de siglo transcurrido de 1974 a 1999. La cifra anterior incluye a 19 países (Alemania, Bélgica, Canadá, China, Dinamarca, Escocia, España, Finlandia, Francia, Hungría, Inglaterra y el País de Gales, Irlanda, Japón, Holanda, Nueva Zelanda, Portugal, la República Checa y Suecia), y las ocho provincias canadienses de Alberta, Columbia Británica, Manitoba, Nueva Brunswick, Newfoundland, Nueva Escocia, Ontario, Quebec y Saskatchewan. Corea mantuvo la situación, eliminando unos exámenes para substituirlos por otros. En sentido contrario, solamente Australia, Grecia y la provincia canadienses de la Isla del Príncipe Eduardo, muestran un decremento en las pruebas en gran escala. (Phelps, 2000: 13-15)

Además de los países mencionados, de los cuales se tiene más información, se sabe que en muchos otros lugares existen sistemas de evaluación de la calidad educativa más o menos consolidados. Destacan los de algunos países del norte de Europa, como Finlandia, Suecia y Noruega, pero también Letonia, Estonia y Li-

tuania; los de algunos países del Asia Oriental, como Corea del Sur y Singapur, cuyo sistema es considerado, al parecer, como el más completo de todos; otros en el Cercano Oriente, en especial Israel, cuyo sistema es muy sólido, y en varios países árabes, donde comienzan a implantarse con apoyo de la UNESCO. En África sobresale la iniciativa de varios países del sur del continente, que a principios de los noventa, y también con apoyo del IIPE de la UNESCO, unieron esfuerzos para desarrollar un sistema de evaluación que incluye la utilización de pruebas aplicadas en gran escala: el *South African Consortium for the Monitoring of Educational Quality*, SACMEQ. Dicho sistema se ha consolidado y extendido a lo largo de la década.

2.3.7. Trabajos internacionales

Aun habiendo sistemas nacionales de evaluación establecidos, la comparación de sus resultados no sería tarea sencilla, dadas las diferencias considerables de los sistemas educativos, en cuanto a estructura, *currículo* y calendarios escolares, además de las diferencias de contenido, grado de dificultad y enfoque de los instrumentos de evaluación mismos. Por ello los trabajos pioneros de la *International Association for the Evaluation of Educational Achievement* (IEA), son tanto más notables (véase Postlethwaite, 1985; de Landsheere, 1994; y Ross, 1992).

Los trabajos de la IEA

A partir de reuniones realizadas a mediados de los años cincuenta en el Instituto de Educación de la UNESCO en Hamburgo, ante la insatisfacción por el uso de tasas de graduación o eficiencia terminal como único indicador de la calidad educativa, en 1958 se planteó la posibilidad de un trabajo de evaluación a nivel internacional, que permitiera obtener resultados estrictamente comparables, mediante la utilización de instrumentos equivalentes en contenido y grado de dificultad. El primer estudio, que tuvo un carácter de piloto, involucrando a 12 países pero con muestras reducidas, se organizó en 1959; los datos se recogieron durante 60-61, fueron procesados en 1961, y se publicaron en 1962. Con la experiencia adquirida se diseñó el primer trabajo en gran escala, sobre Matemáticas, en los 12 países, pero con muestras ma-

yores. La etapa central de recolección de datos tuvo lugar en 1964, y se desarrolló con financiamiento norteamericano.

En 1966 la IEA se constituyó como organización con personalidad jurídica propia, y durante el resto de la década y la de los setenta condujo varios trabajos más: sobre ciencias en 19 países; sobre Comprensión lectora en 15; Literatura en 10; educación cívica en 10; francés e inglés como segunda lengua en 18; Matemáticas, Ciencias e Historia con el entorno del aula en diez países. En los ochenta la IEA llevó a cabo un segundo estudio sobre Matemáticas, en veinte países; también un segundo trabajo sobre Ciencias en 24; y otro sobre Composición escrita en 14. De fines de los ochenta a mediados de los noventa la asociación condujo un trabajo más sobre el uso de computadoras en educación, en 23 países; otro sobre educación preprimaria en 14; uno más sobre Lectoescritura (*reading literacy*) en 31; y el tercer estudio sobre Matemáticas y Ciencias, en el que participaron más de cuarenta países (*Third International Mathematics & Science Study*, TIMSS). Las actividades de la IEA se han fortalecido en los últimos años, dado el creciente interés desarrollado en todos los países por la evaluación. El fortalecimiento se manifiesta de varias maneras:

- Cada vez son más los países que participan en las evaluaciones.
- Los intervalos entre dos aplicaciones sucesivas de una misma prueba se reducen: pasaron diez años entre la primera y la segunda aplicación de la evaluación de Matemáticas y Ciencias (FIMSS & SIMSS), al igual que entre la segunda y la tercera (SIMSS & TIMSS); el TIMSS, en cambio, se aplicó en 1995, luego en 1999, y estuvo en proceso una nueva ronda para 2003, planeándose establecer ciclos similares de cuatro años entre las dos aplicaciones.
- La cantidad de proyectos de evaluación ha crecido, incluyendo algunos innovadores, como los de educación cívica y alfabetización de adultos, además de los convencionales de Matemáticas y Ciencias y de lectoescritura para alumnos de diversos grados.

Muchos investigadores educativos importantes en el nivel internacional han participado en forma destacada en los trabajos de la IEA, incluyendo al sue-

co Torsten Husén, su primer presidente formal; Neville Postlethwaite, responsable del primer estudio, y primera persona que trabajó expresamente para la Asociación; a personas como Fernand Hotyat y Gilbert de Landsheere de Bélgica; Gastón Mialaret de Francia; investigadores de Inglaterra, Holanda, Australia y Nueva Zelanda, como G. Peaker, A. Purves, T. Plomp, J. P. Keeves, además de estudiosos americanos tan destacados como C. Arnold Anderson, Benjamín Bloom, James S. Coleman, Robert L. Thorndike y John Tukey (Husén y Postlethwaite, 1996).

La IEA contribuyó en forma importante a desarrollar la capacidad técnica de investigación educativa en general, y de evaluación en particular, en los países que participaron en los proyectos organizados por ella. En particular el proyecto PISA de la OCDE (véase infra) se ha beneficiado con la experiencia adquirida en los proyectos de la IEA.

El IAEP

Hasta principios de la década de 1990, además de los trabajos de la IEA, hubo otros para construir sistemas de evaluación que permitieran análisis comparativos a nivel internacional, en particular los del *Educational Testing Service* conocidos con el nombre *International Assessment of Educational Progress* (IAEP), a partir del programa *National Assessment of Educational Progress* (NAEP). Sin embargo, estos trabajos no tuvieron continuidad, por lo que la siguiente afirmación, referida a los trabajos de la IEA, parece sostenerse: *En el campo de la evaluación comparativa hay pocas dudas de que, desde sus inicios a principios de los años sesenta, esta organización ha sido la principal fuente de comparaciones confiables entre sistemas educativos* (Goldstein, 1996: 125).

Los trabajos de la OCDE

Desde la década de los ochenta, en el marco de la necesidad de economías cada vez más competitivas en un mundo crecientemente globalizado, y con la conciencia del valor de una buena educación en la nueva sociedad del conocimiento, la Organización para la Cooperación y Desarrollo Económicos (OCDE) promovió un importante proyecto de indicadores educativos al que se ha hecho ya referencia (INES, cfr. su-



pra). A mediados de los noventa, y ante la necesidad de contar con información sobre los resultados de la educación, en la forma de datos comparables sobre el nivel de aprendizaje alcanzado por los alumnos de ciertos grados educativos en los países que la integran, para poder construir indicadores que no se limitaran a los insumos o los datos más elementales, la OCDE decidió emprender el proyecto denominado *Program for International Student Assessment* (PISA).

La preocupación por la calidad educativa, que se ha extendido a todos los países, ha hecho que un número creciente de los pertenecientes a la OCDE hayan solicitado participar en el proyecto PISA, así como en el proyecto de indicadores, INES. El interés internacional que provocó la difusión de los primeros resultados de las evaluaciones del PISA, el 4 de diciembre de 2002, muestra con claridad, la trascendencia de este programa, así como la creciente conciencia de la importancia de la calidad educativa y de su evaluación.

Como proyecto permanente, la planeación de PISA contempla que las evaluaciones se desarrollarán en ciclos separados por intervalos de tres años, en cada uno de los cuales se evaluarán tres grandes áreas de suficiencia de conocimientos y habilidades: Lectura, Matemáticas y Ciencias, dando mayor relevancia a cada una de ellas en las sucesivas aplicaciones. Así en la aplicación del año 2000 el énfasis se puso en el dominio de la Lectura, y las otras dos áreas recibieron una importancia menor; en 2003 el énfasis se pondrá en el dominio de las Matemáticas; en 2006 el foco estará puesto en las ciencias; y a partir de 2009 el ciclo se repetirá.

Las pruebas de PISA se distinguen también por la utilización de las metodologías más avanzadas tanto para el diseño de las pruebas, como para el análisis de sus resultados. Se desarrollaron complejas especificaciones que combinan campos de aplicación con niveles de desarrollo cognitivo. Peculiar también es la decisión de aplicar las pruebas a jóvenes de 15 años, que es la edad más temprana en que termina la obligatoriedad escolar en los países de la OCDE, incluyendo a todos los que estén en el sistema educativo, independientemente del grado en que se encuentren, y descartando a quienes hayan salido ya del sistema. (Cfr. PISA, 2000 y 2001)

Los trabajos latinoamericanos

Como antecedente puede mencionarse el proyecto Estudios Conjuntos sobre la Integración Económica Latinoamericana (ECIEL), coordinado por Claudio de Moura Castro, quien en los años setenta estudió el nivel de preparación de la fuerza de trabajo de varios países de América Latina utilizando pruebas de la IEA (véase de Moura Castro, 1980; Muñoz Izquierdo 1976; Wolff, 1998).

En fechas más recientes, particularmente de mediados de los noventa en adelante, la conciencia de la importancia de la evaluación educativa en los países de América Latina, ha hecho que se desarrollen iniciativas multinacionales en este campo, entre las cuales pueden mencionarse el Programa Iberoamericano de Evaluación de la Calidad Educativa (PIECE), de la Organización de Estados Iberoamericanos; el Proyecto de Establecimiento de Estándares para la Educación Primaria de la OEI y la Coordinación Educativa y Cultural Centroamericana; el Programa de Promoción de la Reforma Educativa en América Latina y el Caribe, PREAL; el Proyecto Regional de Indicadores Educativos, derivado de la segunda Cumbre de las Américas, coordinado por el Ministerio de Educación de Chile; y el Forum Hemisférico de Evaluación Educativa, derivado de la II Cumbre, y coordinado a su vez por el INEP de Brasil.

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

Los Sistemas Nacionales de Medición y Evaluación de la Calidad de la Evaluación que, con diversas denominaciones y grados de consolidación, existen en 16 países de la región, participan en las actividades del LLECE; si bien no todos lo han hecho en la misma medida y en todos los aspectos: Argentina, Bolivia, Brasil, Chile, Colombia, Costa Rica, Cuba, Ecuador, El Salvador, Guatemala, Honduras, México, Panamá, Paraguay, República Dominicana y Venezuela.

En 1997 se llevó a cabo la primera aplicación de pruebas de Lenguaje, Matemáticas y factores asociados, a muestras de alumnos de 3° y 4° grado de educación básica; a fines de 1998 se publicó el primer informe de resultados, y en el año 2000 el segundo. La publicación de un informe técnico detallado se previó para el primer semestre de 2002. (Wolff, 1998; LLECE, s/f, 2000)

2.4. La evaluación y la medición educativa en México

México no se excluye a la creciente tendencia mundial en pro de la evaluación educativa; de hecho, como se ha apuntado, el país se anticipó a las demás naciones latinoamericanas en este terreno, si bien lo hizo de manera limitada, por lo que fue hasta mediados de los años noventa, como en el resto del subcontinente, cuando la tendencia comenzó a consolidarse. En el último inciso de este apartado, se resume la experiencia mexicana en el campo de la evaluación educativa. Como en otros lugares, en México la evaluación del aprendizaje de los alumnos es realizada por el maestro en forma individual. La aplicación de procedimientos sistemáticos, equivalentes, o iguales, a grandes números de estudiantes, no ha sido usual en nuestro medio. Si bien hay antecedentes más antiguos, la utilización de pruebas en gran escala para la evaluación del Sistema Educativo Nacional comienza apenas, con fuertes limitaciones, en la década de los años setenta, cuando la Secretaría de Educación Pública comenzó a aplicar pruebas estructuradas de aprendizaje para fundamentar las decisiones sobre el ingreso de los alumnos a secundaria, y luego, a muestras nacionales de niños de primaria. Hasta los noventa, con la ampliación de estas acciones de la SEP para la educación básica, y la creación del Ceneval para la media superior y la superior, cuando este tipo de evaluaciones comenzó a extenderse. La historia de estas actividades en México se resume en los siguientes puntos.

2.4.1. Antecedentes de la evaluación, de 1936 a 1958

Oficialmente, el antecedente más remoto de la investigación y la medición educativa en México lo constituye el Instituto Nacional de Psicopedagogía,

creado en 1936 por decreto de Lázaro Cárdenas. Sin embargo, esta institución no desarrolló una actividad significativa durante los períodos presidenciales del propio Cárdenas (1934-1940), y sus dos sucesores, Manuel Avila Camacho (1940-1946) y Miguel Alemán (1946-1952), al grado de que al comenzar el sexenio de Adolfo Ruiz Cortines estuvo en riesgo de desaparecer. Conviene señalar que en 1935 se había establecido un Instituto Médico Pedagógico, surgido con la tarea de apoyar la educación de los niños que se clasificaban como débiles mentales que, según los estudios del Tribunal de Menores de la época, representaban el 67% de los menores infractores (DGESIC, ca. 1959: 53, 78).

Aunque no es fácil contar con información sobre los inicios del INP, hay indicios de una actividad interesante que, de haber tenido continuidad, podría haber conducido a un desarrollo más inmediato de la investigación psicométrica en el país. Una obra de la época informa que, ante la insatisfacción que provocada por contar solamente con una prueba de tipo verbal para evaluar la inteligencia (la estandarización Santamarina de la Prueba Individual Binet-Simon). Desde finales de 1936 se decidió llenar la laguna mediante *la estandarización de una prueba de ejecución que permita medir la habilidad mental a través del movimiento, que es otra de sus expresiones funcionales*. Para ese propósito se seleccionó la Prueba Individual de Ejecución de Cubos y Dibujos de S. C. Kohs, cuya correlación con la Revisión Stanford de la Binet-Simon, *considerada en el mundo entero como la mejor escala de medida de la inteligencia general* es reportada precisamente por el mexicano, con base en el estudio monográfico del propio Kohs: *coeficiente de correlación entre la edad mental que determina la Prueba Binet y la edad mental determinada por Kohs: + 0.82 con un error probable de 0.01 con una muestra de 366 casos* (López, et al., 1938: 3-4).

El trabajo de estandarización fue hecho por el profesor Matías López, del Servicio de Psicometría, con ayuda de las profesoras Etelvina Acosta, Guadalupe Díaz y Ma. Eulalia Benavides. El informe, que utiliza una metodología simple, pero cuidadosamente seguida, muestra también un conocimiento de los autores más importantes del campo, ya que se cita pertinentemente a Wundt, Neuman y James, pero también a Bain, Royce y Thorndike, así como a autores menos conocidos, como Hobhouse, Ti-



chener, Maudsley, Ladd y Woodworth. (López, *et al.*, 1938: 11-27).

Al comenzar el sexenio 1952-1958, la gravedad de la problemática educativa, en un momento cuando el crecimiento demográfico se volvía explosivo, hacía necesarios los trabajos rigurosos de investigación educativa, por lo que se decidió apoyar al Instituto en la tarea de desarrollar ese tipo de estudios, en colaboración con el nuevo Consejo Nacional Técnico de la Educación. Se incrementó el presupuesto y se crearon plazas técnicas y administrativas. Así, el total del personal adscrito aumentó de 68 personas en el sexenio 1946-1952 a 95 en el 1952-1958. Más significativo que el incremento numérico lo es el tipo de personal que los totales anteriores incluyen. Los trabajadores del Instituto entre 1946 y 1952, no eran pocos, pero su composición revela escasa competencia profesional para realizar investigaciones educativas, ya que estaba formado por siete médicos, veinte maestros, nueve trabajadoras sociales, 19 técnicos no titulados y 13 administrativos. En el sexenio 1952-1958 se contrataron dos médicos, 15 maestros, una trabajadora social y tres técnicos, reduciéndose el personal administrativo de 13 a 7 personas; se contrataron además 12 profesores con formación en orientación, perfil no existente aún.

El informe de labores del sexenio 1952-1958 reporta que en ese período los laboratorios del Instituto atendieron a mil 800 personas en exámenes clínicos; 4 mil en exámenes médicos; 2 mil 400 en exámenes de agudeza visual y 552 para tratamientos de ortolalia. Además de hicieron 3 mil 600 estudios socioeconómicos y se aplicaron *pruebas mentales* a niños de primaria y secundaria. En este último aspecto se reportan 160 mil pruebas en 1953; 135 mil en 1954; 140 mil en 1955; y 160 mil en 1956.

Cada uno de estos años las pruebas se aplicaron a un número de escuelas que varió de 122 a 140; se cubrieron todos los grados, con un número mayor en primero de primaria (de 20 mil a 26 mil) y cifras algo menores (de diez mil a veinte mil) en los demás, hasta tercero de secundaria (DGESIC, ca. 1959: 54-59).

La orientación del modesto trabajo de investigación que se realizaba en el INP puede inferirse de los datos anteriores: una vertiente fundamental era de tipo médico; la otra incluía la aplicación de pruebas, pero sin indicio alguno, de que se utilizaran los

avances hechos para mediados de siglo por la psicometría norteamericana. Esta impresión se confirma revisando otros elementos relacionados con la investigación educativa que recibieron impulso durante el sexenio 1952-1958.

- Se fortaleció la Escuela Normal de Especialización, fundada en 1943. La Escuela preparaba maestros en cuatro especialidades dedicadas a la atención de niños ciegos, sordos, lisiados y *normales mentales, inadaptados y menores infractores*. Además, la Escuela contaba con trabajadoras sociales que hacían estudios de *paidografía social* y tenía anexo el Instituto Médico Pedagógico 1935, que tenía servicios médicos y dentales, servicio social, un Departamento Pedagógico, un Gabinete de Material Didáctico y un Gabinete de Psicotécnica. En éste se aplicaban pruebas de inteligencia según las cuales el cociente intelectual de los niños evaluados, se situaba en promedio alrededor de un valor de setenta.
- Se dio nuevo impulso al Museo Pedagógico Nacional, fundado en 1944 y concebido como *institución de carácter científico y técnico, el cual tenía por misión la investigación científica de la evolución histórica del movimiento educativo del país y del extranjero, así como la concentración de datos de la educación pública en general y la información documental y objetiva en todos los aspectos de la educación*. El número de plazas de planta de esta institución aumentó de 20 a 35 (DGESIC, ca. 1959: 103-108).
- Se apoyó a la Clínica de Conducta, fundada en 1948 que, según el reporte, había atendido a 3 mil escolares entre 1948 y 1952 y recibió a 7 mil en el lapso 1953 a 1958. Del total de unos 10 mil casos en la década 800 corresponden al rubro de *psicometría*. (DGESIC, ca. 1959: 119-121)

2.4.2. Investigación y evaluación en los sesenta y setenta, sobre todo en educación superior

Ante los modestos esfuerzos anteriores, la década de 1960 constituyó, en opinión general, la época del inicio real de la investigación educativa en México, con la fundación del Centro de Estudios Educativos por Pablo Latapí, en 1963. Sin embargo, el nuevo Centro como harían luego otras instituciones interesadas en estos temas, atendió otras dimensiones de la proble-

mática educativa, en especial las de tipo social, desde la perspectiva de disciplinas como la sociología y la economía primero, y la antropología después.

El desarrollo de los aspectos psicométricos se dio más bien en el Colegio y luego Facultad de Psicología de la UNAM. En esta institución el interés educativo fue también secundario, frente al clínico y otros campos de aplicación de la Psicología. Por todo lo anterior, no resulta sorprendente el que ninguna de las instituciones mencionadas haya cultivado sistemática y profesionalmente la psicometría. Sin embargo, y aunque en forma que podría llamarse artesanal, la elaboración de pruebas de aprendizaje de tipo objetivo, comenzó desde los años sesenta y principios de los setenta, en general, con propósitos de selección, aunque no exclusivamente. Posiblemente la experiencia más antigua en ese sentido sea la de la Facultad de Medicina de la UNAM, que a mediados de la década de 1960 desarrolló un banco de reactivos de opción múltiple, para el examen de titulación de la carrera de Médico cirujano, calificado por computadora.

Fue importante también la experiencia de la Facultad de Ingeniería de la UNAM donde, a partir de 1976, y además de pruebas de diagnóstico para alumnos de nuevo ingreso, se incursionó en forma imaginativa en el campo de la evaluación: los exámenes departamentales de las materias derivadas de Matemáticas se hacían por computadora, con modificaciones propias al modelo de la teoría clásica, para la toma de decisiones respecto a la calidad de los reactivos; se emitían dictámenes de diverso grado de agregación para los maestros, el coordinador de la carrera, el director de la división y la dirección de la Facultad, además de informes de retroalimentación para las escuelas de origen de los aspirantes. Los informes eran especialmente cuidadosos, con recomendaciones personalizadas al destinatario sobre aspectos a mejorar, incluyendo datos y referencias precisas de obras a consultar; se desarrolló un modelo de banco de reactivos con un concepto de apoyo a los profesores, como un medio para que mejoraran la calidad de sus propias evaluaciones; se desarrolló también un programa de calificación que permitía manejar, no sólo reactivos de opción múltiple, sino también de falso verdadero, jerarquización, relación de columnas e inclusive de complementación y respuesta abierta breve. (Información personal del Ing. Agustín Tristán).

Sin embargo, seguramente la experiencia más importante fue la relativa a los exámenes de selección para ingreso a la UNAM, tanto en el nivel de bachillerato como en el de licenciatura, procesos que se volvieron complejos y delicados, cuando el número de aspirantes se multiplicó y los lugares disponibles fueron cada vez más insuficientes.

Para atender esta problemática, la UNAM estableció en 1963 un área especializada a cuyo frente estuvo muchos años la Dra. Ema Dolujanof, que debe ser considerada seguramente la persona más importante en este terreno en la etapa a que se refiere este apartado. Sin un acercamiento sofisticado, pero con gran cuidado y competencia, la Dra. Dolujanof logró desarrollar pruebas que, sometidas, *a posteriori*, a evaluaciones psicométricas, mostraron niveles de confiabilidad excelentes e, inclusive, grados de comparabilidad de la versión de un año a las siguientes no muy inferiores a los que se alcanzan con las modernas técnicas de igualación o equiparación (*equating*). Los trabajos de esta área de la UNAM, continuados luego en el marco de la Dirección General de Planeación, fueron la base para la creación del Ceneval en 1994 (comunicación personal del Maestro Rafael Vidal).

Otras instituciones de educación superior tanto privadas, como la Universidad Iberoamericana, como públicas, como la Autónoma de Aguascalientes, desarrollaron sus propias pruebas de ingreso durante los años setenta, también en la forma que llamamos artesanal, sin controlar sistemáticamente las propiedades psicométricas de las pruebas y, en consecuencia, sin poder hacer análisis comparativos o longitudinales. El Instituto Tecnológico y de Estudios Superiores de Monterrey y la Universidad de las Américas, comenzaron a utilizar los servicios de la Oficina establecida por el College Board en Puerto Rico en 1963, contratando la aplicación de la Prueba de Aptitud Académica (PAA), equivalente al SAT, para propósitos de selección para el ingreso al nivel de licenciatura.

2.4.3. La evaluación educativa en la SEP, en los setenta y los ochenta

Desde principios de la década de 1970, la SEP comenzó a realizar actividades de evaluación, que llevaron a la creación de un área especial para ello, la Dirección General de Evaluación (DGE) de la Sub-



secretaría de Planeación y Coordinación, actualmente Dirección General de Políticas Educativas (DGP). Las primeras pruebas que se aplicaron en gran escala, en 1972, se utilizaban para decidir la admisión de alumnos a la secundaria. Luego, en el sexenio 1976-1982 se hicieron las primeras evaluaciones de primaria mediante la aplicación de pruebas a muestras nacionales, con el proyecto llamado *Evaluación del rendimiento académico de los alumnos de 4° y 5° de educación primaria*. Desde entonces la DGE ha aplicado pruebas de tal tipo a muestras nacionales de alumnos. Las preocupaciones políticas, sin embargo, impidieron que esas actividades se consolidaran; los resultados no se difundían y, hasta donde puede saberse, no se aprovechaban para retroalimentar a los tomadores de decisiones (véase Dirección General de Evaluación, 1997b). La no consolidación de las actividades de evaluación se manifestaba en las deficiencias técnicas de los trabajos que se hacían sin aplicar las técnicas psicométricas y estadísticas necesarias para garantizar la equiparabilidad de las versiones y la estabilidad de las escalas utilizadas; el sustento que tan rudimentarios instrumentos podía ofrecer a los tomadores de decisiones, era en consecuencia, muy precario, dado que no existían parámetros de referencia bien definidos, y no era posible hacer comparaciones consistentes entre subsistemas estatales o de otro tipo, y menos aún análisis longitudinales.

Según la información del capítulo sobre las actividades de la Dirección General de Evaluación de la *Memoria 1995-2000*, de la SEP, hasta principios de los años noventa la experiencia relativa a evaluaciones en gran escala incluía los exámenes de ingreso a la educación normal, los de Diagnóstico para Alumnos de Nuevo Ingreso a Educación Secundaria (IDANIS) que se aplicaban en sólo tres entidades, y el Estudio Integral de la Evaluación Preescolar, Primaria y Secundaria, aplicado en nueve entidades a 190 docentes (Velázquez, 2000: 647).

2.4.4. Los años noventa y el inicio del siglo XXI

El desarrollo cualitativo de la evaluación educativa en México, como en la mayor parte de los países del mundo y, en particular, de América Latina, se da hasta la última década del siglo XX.

En lo que se refiere a los niveles más altos del sistema educativo, el paso más importante al respecto fue la fundación del Centro Nacional para la Evaluación de la Educación Superior (Ceneval), en 1994, precedido por el lanzamiento del Examen de Conocimientos y Habilidades Básicas (Excohba) de la Universidad Autónoma de Baja California (1990). De la misma fecha data la incursión masiva en el mercado mexicano de las pruebas de aptitud académica en Español de la oficina de Puerto Rico del College Board.

El Ceneval aplicó 365 mil pruebas en su primer ciclo anual de trabajo, de mayo de 1994 a abril de 1995; las pruebas se limitaban básicamente a las de ingreso a educación media superior y a educación superior (EXANI I y EXANI II). En el ciclo 1995-1996 la cifra ascendió a 492 mil 618; pasó a 561 mil 018 en 1996-1997; 973 mil 770 en 1997-1998; y 1 millón 222 mil 303 en el año civil de 1999. Si bien las pruebas del EXANI I y el EXANI II siguen siendo, con mucho, las más numerosas, el Ceneval maneja más de 25 Exámenes de Egreso de Licenciatura (EGEL), y prepara varios más, que se aplican a números crecientes de egresados.

Por su parte, el *College Board*, que en 1989 aplicaba anualmente unas 25 mil pruebas de aptitud para ingreso a licenciatura en toda América Latina, incluyendo unas cuantas instituciones mexicanas de educación superior, todas privadas, comenzó en 1990 a incursionar en el mercado de las universidades públicas, con la Autónoma de Aguascalientes, y en 1999 aplicaba más de 50 mil pruebas de aptitud al año sólo en México, en instituciones públicas tan grandes como la Universidad de Guadalajara y la de Puebla.

Además del Ceneval, desde el sexenio 1989-1994, se crearon otros organismos de evaluación de instituciones de educación superior, que se fueron consolidando en la segunda mitad de los noventa: en 1991, los Comités Interinstitucionales de Evaluación de la Educación Superior (CIEES); en diversos momentos agencias acreditadoras especializadas; y, a fines del 2000, un organismo para coordinar lo relativo a la acreditación: el Consejo para la Acreditación de la Educación Superior, Copaes.

En lo concerniente a educación básica ocurrieron cambios importantes a fines del sexenio 89-94, sobresaliendo la inclusión de los factores de rendimiento

escolar y profesional en el esquema del programa de estímulos de Carrera Magisterial, lo que trajo consigo la necesidad de aplicar cada año pruebas de conocimientos a números considerables de maestros (más de 500 mil incrementándose el número constantemente) y a sus alumnos (cerca de 4 millones 500 mil también en incremento). Con lo anterior el volumen de las acciones evaluativas de la DGE creció substancialmente, con los problemas de calidad que tal acrecentamiento hacía inevitables, con la falta de una tradición psicométrica. No podía hablarse de un acercamiento integral; se desatendían muchos aspectos (Velázquez, 2000: 647).

En el sexenio 1995-2000 la situación comenzó a modificarse paulatinamente: algunas entidades de la República, como Aguascalientes, Guanajuato y Nuevo León, comenzaron a desarrollar sistemas de evaluación propios (Cfr. Martínez Rizo, 1993 y 1991; SEG, 1995; UIE, 1996). Por su parte, la SEP, además de apoyar la creación de áreas de evaluación en los estados, se propuso desarrollar un Sistema Nacional de Evaluación Educativa, SNEE (véase Velázquez, 2000; también DGE 1996a, b y c; 1997a y b; Martínez Rodríguez, 2000) y dio pasos importantes en esa dirección, incluyendo acciones derivadas de la participación de México en proyectos internacionales de evaluación (cfr. Tiana Ferrer, 2000). A partir del capítulo ya mencionado de la *Memoria* (Velázquez 2000: 643-682), las principales actividades de la DGE en este campo en el sexenio 1995-2000 se resumen como sigue:

- En colaboración con el Ceneval, a partir del ciclo 96-97, se unificaron los procesos de selección de alumnos para las instituciones de nivel medio superior de la Zona Metropolitana de la Ciudad de México, constituyéndose la Comisión Metropolitana de Instituciones Públicas de Educación Superior (Comipems), bajo cuya coordinación se aplican anualmente alrededor de 250 mil exámenes. Sin duda, la novedad y las dimensiones de la experiencia explicaron sin duda la presencia de diversas fallas en la primera aplicación, que se han corregido progresivamente. (Velázquez, 2000: 650-655)
- Comenzando 1991, el gobierno federal, con apoyo del Banco Mundial, promovió un importante programa educativo compensatorio en gran es-

cala, en beneficio de cuatro de las entidades más pobres, conocido con el nombre de Programa para Abatir el Rezago Educativo (PARE), cuya evaluación incluyó la aplicación de pruebas de aprendizaje durante cinco ciclos escolares consecutivos, hasta 1995. A partir de esa fecha, y con base en la experiencia obtenida, la DGE amplió considerablemente la gama de actividades evaluativas, en un gran esfuerzo que, en conjunto, no es otra cosa sino el llamado Sistema Nacional de Evaluación Educativa (SNEE). Dichas actividades incluyeron:

- El diseño del componente de evaluación del programa compensatorio que siguió al PARE, el Programa para Abatir el Rezago en Educación Básica (PAREB).
- El proyecto denominado Estudio de Evaluación de la Educación Primaria (EVEP), con el que se buscó extender la aplicación de pruebas similares a las de los programas compensatorios a todas las entidades, en forma permanente.
- Diversos esfuerzos de tipo cualitativo, junto con actividades para desarrollar la capacidad de evaluación en el aula de los maestros.
- El impulso a la creación o consolidación de Áreas Estatales de Evaluación (AEE) en todas las entidades de México, para incrementar la capacidad de utilizar los resultados de las evaluaciones hechas por la DGE y de hacer otras, lo que implicó diversas acciones de equipamiento y capacitación para el personal.
- La constitución de un mecanismo de colaboración entre los responsables de dichas AEE, el llamado Comité Técnico Interestatal de Evaluación, CTIE. (Velázquez, 2000: 655-662)
- Continuó la aplicación de pruebas para evaluar el factor de preparación profesional de los maestros que participan en el programa de Carrera Magisterial, así como el del aprovechamiento escolar de sus alumnos, con mejoras en lo relativo a la consolidación de una base de datos nacional al respecto, el diseño de instrumentos, la logística para el levantamiento de los datos, la lectura de los mismos y la emisión de los resultados. Por las dimensio-



nes de esta actividad, lo anterior representó el mayor esfuerzo de la DGE y ha traído consigo aprendizajes importantes. (Velázquez, 2000: 662-669)

- Aunque menor en términos cuantitativos, cualitativamente es más importante el trabajo de definición de estándares curriculares nacionales, a partir de los cambios en planes y programas de estudio de 1993, junto con el trabajo de construcción de instrumentos de evaluación del aprendizaje alcanzado por los alumnos en relación con esos estándares. Aunque aún no se habían publicado resultados, seguramente este esfuerzo es el más avanzado técnicamente en el terreno de evaluación de los realizados hasta esa época por la DGE; la primera medición que utilizó los nuevos instrumentos se hizo al inicio del ciclo 98-99, la segunda al inicio del ciclo 1999-2000, y la tercera a fines del mismo. Se aplicaron también instrumentos para obtener información sobre el contexto de la enseñanza, con propósitos explicativos. (Velázquez, 2000: 669-670)
- Continuó aplicándose el Instrumento de Diagnóstico para Aspirantes de Nuevo Ingreso a Secundaria, IDANIS, así como el Instrumento de Diagnóstico y Clasificación para Ingreso a la Educación Normal (IDCIEN). (Velázquez, 2000: 670-671)
- Se impulsó un trabajo de evaluación de escuelas con acercamiento cualitativo, buscando estudiar los factores asociados al aprovechamiento escolar, lo que supuso una importante gama de actividades de capacitación en investigación cualitativa para el personal de las Áreas Estatales de Evaluación de cuatro entidades, seguidas por otras diez y luego por otras 15. (Velázquez, 2000: 673-674)
- Con el estímulo que representó el ingreso de México a la OCDE, el país comenzó a participar en proyectos internacionales de evaluación educativa: el TIMSS de la IEA, del que se retiró al final para evitar que se publicaran los resultados del país, que eran inferiores a los de otros participantes; el Laboratorio Latinoamericano de Evaluación de la Calidad Educativa de la OREALC; y más recientemente el proyecto PISA, de

la OCDE. El elevado nivel técnico de las evaluaciones internacionales representa a la vez un reto importante y una excelente oportunidad de aprendizaje. (Velázquez, 2000: 675-678)

- En paralelo a lo anterior, la DGE desarrolló una labor editorial significativa, publicando diversos textos para capacitación e informes de resultados, si bien estos han sido difundidos de manera reducida y selectiva. Por otro lado, se reporta que no ha habido avances comparables en el terreno de la evaluación de políticas y sistemas educativos. Se informa también sobre algunas dificultades que han impedido un desarrollo más vigoroso de las actividades de evaluación, en especial la elevada rotación de personal, derivada de los bajos salarios, que dificulta la profesionalización, y, en general, la gran escasez de recursos. (Velázquez, 2000: 671-673 y 681-682)

En los últimos años, a través del SNEE se han desarrollado los primeros esfuerzos importantes para establecer las bases de una cultura de la evaluación en un contexto nacional que, en general, había carecido de ella. Los resultados de las evaluaciones de Carrera Magisterial y otras fueron entre autoridades educativas estatales y, en la medida en que éstas se han interesado y comprometido con la iniciativa, han llegado a las escuelas. Algunos proyectos educativos, tales como el de Gestión Escolar en la Escuela Primaria, a cargo de la Dirección General de Investigación Educativa de la SEP, han hecho uso extensivo de los materiales disponibles y a su vez han contribuido al aprovechamiento de los mismos para mejorar la calidad de la educación.

Sin embargo, es claro que es necesario seguir impulsando una mayor conciencia entre la sociedad en general y el sector educativo en particular, de los beneficios que puede aportar la evaluación. Resulta igualmente necesario seguir desarrollando las capacidades y competencias requeridas para la práctica de la medición del logro académico, así como su interpretación correcta y aprovechamiento en las entidades federativas, debido a la escasa tradición previa existente en México en este campo, hasta ahora el nivel técnico del sistema sigue estando por debajo de los estándares de calidad que los avances internacionales hacen posibles y deseables. De igual forma, es necesario adecuar la oferta de servicios y aplicacio-



nes a las necesidades de información de las escuelas y los sistemas educativos.

Para alcanzar niveles técnicos elevados en el terreno de la evaluación, y más allá de los esfuerzos por desarrollar la capacidad de las entidades federativas al respecto, parece también que la estructura actual del SNEE mejora sustancialmente. Además de la autoridad educativa federal y de las autoridades estatales (SEP y secretarías estatales), cuya responsabilidad en lo relativo a evaluación no puede esclarecerse, es evidente la necesidad de uno o más órganos autónomos, con recursos técnicos y personalidad jurídica tales que permitan la realización de todos los aspectos técnicos de la evaluación con la calidad necesaria. La Dirección General de Evaluación coincide con esta opinión, según se expresa en el capítulo multicitado de la *Memoria* (Velázquez, 2000: 673).

En el terreno metodológico, un avance fundamental lo constituye la aparición de instancias profesionales de elaboración de pruebas (Ceneval, DGE de la SEP) que están haciendo instrumentos que se acercan poco a poco a los estándares de calidad internacionales. Sin embargo, la falta de tradición psicométrica a que se ha hecho referencia hace que aún ahora haya limitantes serias, no todas las cuales pueden imputarse a los responsables de la elaboración de las pruebas.

De parte de éstos, ciertamente, debe señalarse que los avances más importantes de la psicometría de la segunda mitad del siglo XX, siguen siendo poco conocidos. Aunque ha comenzado a usarse la Teoría de la Respuesta a los Ítems, las pruebas siguen siendo fundamentalmente de opción múltiple, en tanto que las de respuesta construida, especialmente las de ejecución, son inexistentes, lo mismo que las pruebas computarizadas adaptativas o las de papel y lápiz, que utilicen muestreo matricial para aplicaciones en gran escala, como las que debe realizar la SEP a nivel nacional.

Además de lo anterior, hay limitaciones importantes que son responsabilidad de otras instancias: no pocas instituciones de educación superior, por ejemplo, pese a las recomendaciones técnicas en contrario, utilizan los resultados en una prueba estandarizada de ingreso como único elemento de juicio para tan delicada decisión, sin incorporar un elemento como el promedio obtenido por cada sustentante en el ni-

vel escolar anterior, que mejoraría significativamente la calidad de la decisión, el cual es bastante fácil de obtener, inclusive haciendo correcciones estadísticas para reducir la diversidad de criterios para asignar calificaciones en las diferentes escuelas de procedencia.

Otra limitante no técnica muy importante es el manejo de los resultados de las evaluaciones en gran escala, las cuales algunas autoridades parecen considerar como *secreto de estado*, en lugar de darles difusión, cuidando la privacidad de las personas involucradas, difusión que es fundamental para los propósitos de mejoramiento, en relación con el concepto de rendimientto de cuentas (*accountability*).

Como otra limitación no imputable a las instancias dedicadas a la elaboración de pruebas, y que en alguna medida explica, inclusive las de ellas, debe mencionarse el escaso desarrollo de la psicometría y de la investigación educativa en general, y la evaluativa en especial, en las instituciones de educación superior e investigación. Con excepciones como las que constituyen el Instituto de Investigación y Desarrollo Educativo de la Universidad Autónoma de Baja California (UABC), o el Departamento de Educación de la Universidad Autónoma de Aguascalientes (UAA), pocas instituciones universitarias han cultivado estos campos, lo que explica la cantidad prácticamente nula de especialistas en evaluación que se forma en el país.

Si bien es verdad que las autoridades educativas federales han mantenido los resultados de las evaluaciones realizadas como materia confidencial, también lo es que no hay estudiosos que soliciten acceso a esos resultados para hacer análisis propios; tampoco trabajos independientes de evaluación, con pocas excepciones a las cuales se ha hecho referencia, como los mencionados de la UABC y la UAA. Otra excepción digna de mención, es un trabajo de Palafox *et al.*, publicado en la *Comparative Education Review* (1994), que es seguramente uno de los primeros trabajos de su tipo.

2.5. Conclusión

Para terminar, se debe señalar que las implicaciones del uso de pruebas en gran escala y, en general, de las técnicas de medición en psicología y educación,

han propiciado que se desarrollen estándares profesionales para normar el trabajo de los especialistas de ese campo; tales estándares son, a la vez, criterios de calidad que ayudan a los usuarios, en especial autoridades educativas estatales y distritales, a seleccionar unas pruebas con preferencia a otras y utilizar apropiadamente sus resultados. Uno de los esfuerzos más recientes de ese tipo apareció a fines de 1999. Versión de la última interpretación de los *Standards for educational and psychological testing*, preparada conjuntamente por la *American Educational Research Association*; la *American Psychological Association* y el *National Council on Measurement in Education* (AERA-APA-NCME, 1999). También la IEA difundió sus propios estándares, con el título *Technical Standards for IEA Studies*. (Véase Martin, Ed., 1999).

En el prefacio de la obra de Estándares de AERA-APA-NCME se señala que tiene como antecedentes cinco documentos similares, publicados previamente por una de las organizaciones patrocinadoras, o por las tres: en 1954 la APA publicó el primero: *Technical Recommendations for Psychological Tests and Diagnostic Techniques*; un año después, en 1955, la AERA y el NCME (entonces NCMUE), elaboraron el segundo, *Technical Recommendations for Achievement Tests*; en 1966 las tres organizaciones unieron esfuerzos y prepararon un tercer documento, que remplazó a los dos anteriores: fue la primera versión de *Standards for Educational and Psychological Tests and Manuals*; los textos cuarto y quinto fueron las versiones de 1974 y de 1985 de tal texto, con un nombre ligeramente distinto: *Standards for Educational and Psychological Tests*, en 1974, y *Standards for Educational and Psychological Testing*, en 1985 (AERA-APA-NCME, 1999, p. v).

La diferencia entre la versión de los estándares de AERA-APA-NCME publicada a fines de 1999 y las anteriores es notable. La versión previa se refería todavía en forma exclusiva a pruebas más convencionales, de opción múltiple, y se basaba en la teoría clásica de las pruebas, ignorando los avances que ya habían tenido lugar mucho antes no sólo de la última reedición de dicha versión previa, en 1997, sino incluso de su edición de 1985. Lo anterior parece mostrar la lentitud con que los organismos dedicados a la elaboración de pruebas, presionados seguramente por lo pesado de la tarea, incorporan los desarrollos teóricos y técnicos. Por eso es una sorpresa ver que los

nuevos estándares consideran expresamente las nuevas teorías psicométricas y las formas de pruebas novedosas, como las de ensayo y ejecución, portafolios, pruebas adaptativas por computadora, aspectos muy finos relativos al sesgo, muestreo matricial, etcétera.

El desfase cronológico de varias décadas que se para los avances más tempranos en este campo en los Estados Unidos, de los más recientes en otros países, como México, hace necesario que en nuestro país, y en los que se encuentren en situación análoga, se analice la situación de quienes han incursionado antes en este tipo de actividades, no para imitarlos ciegamente, pero sí para aprovechar su experiencia en forma crítica. Esto es tanto más necesario cuanto que parece haber una notable similitud de las situaciones que se comparan, a la distancia de algunas décadas. El conocimiento de experiencias tempranas podrá, seguramente, ahorrar errores y permitir avanzar con rapidez para estar en niveles de calidad comparables a los de los países con más tiempo de trabajo en el campo. El paralelismo de algunos rasgos de la historia del *College Board* y la del Ceneval, por ejemplo, es notable. En el prefacio de la edición de 1984 del *Technical Handbook for the SAT* su principal autor, retomando una idea del historiador del *CollegeBoard*, Claude Fuess, subraya que la historia del organismo “no es la de una agencia productora de pruebas que trata de vender sus servicios a las universidades, sino la de una asociación de *colleges* los cuales se unen con el propósito común de facilitar el paso de los alumnos de las escuelas medias superiores a la universidad”. (Donlon, 1984: XIX).

Al resumir la historia de las evaluaciones en gran escala, se pudo mostrar que los problemas que enfrentan las instituciones dedicadas a esa tarea en México no son inéditas, sino que se han presentado en otros lugares y hay soluciones bien probadas ya para algunos de ellos, en tanto hay otros retos todavía por atenderse. Entre estos destaca el que se refiere a la incorporación necesaria de los avances de las ciencias cognitivas, desarrollados también en la segunda mitad del siglo XX, pero con escaso contacto con la psicometría, al grado de que la opinión expresada por un especialista sea compartida ampliamente: *Es sólo una leve exageración describir la teoría de los tests que domina hoy el campo de la medición educativa, como la aplicación de la estadística del siglo XX a la psicología del siglo XIX*, escribe Robert



J. Mislevy al inicio de su contribución a la importante obra *Test Theory for a New Generation of Tests* (Mislevy, 1993, en Frederiksen *et al.* 1993, 19)

Las tareas de elaborar, aplicar y procesar pruebas de alta calidad para la evaluación de los logros de los alumnos, en relación con los objetivos de diversas áreas del currículo, y que se apegue a los estándares del campo es, de alta complejidad técnica. La experiencia de maestros, directores y supervisores es un insumo de gran valor, pero no basta: es indispensable la presencia de personal altamente calificado para la realización de tan especializadas actividades.

Keeves enumera una serie de competencias especializadas que deberán combinarse adecuadamente para lograr la interrelación adecuada: para elaborar pruebas de alta calidad se requiere del concurso de especialistas en las áreas de conocimiento a evaluar pero, además, en las bases psicopedagógicas del aprendizaje; en medición educativa, psicometría y estadística; en realización de encuestas e investigación educativa en general; en diseño gráfico, publicación y edición; en computación, comunicación, administración y contabilidad; la integración del personal que tenga tan variadas capacidades y la organización de un centro en donde se combinen, además, de considerables capacidades de negociación y liderazgo. (1994: 99)

En algunos casos puede existir la posibilidad de aprovechar instrumentos desarrollados en otro país, pero diversos obstáculos (lingüísticos, culturales, curriculares, entre otros) harán necesario en muchas ocasiones que los países grandes, o grupos de países

pequeños los cuales compartan una misma lengua, cuenten con capacidad propia para este tipo de trabajos.

La ubicación organizacional de una unidad de elaboración de pruebas puede ser muy variada. En las experiencias analizadas tenemos el caso de una entidad privada (ETS), el de un organismo público autónomo dentro del sector educativo (CITO e INCE) o el de una dependencia del Ministerio de Educación (DEP en Francia) y, seguramente, otras variantes. En América Latina parece haber sólo unidades dependientes del ministerio respectivo, con diversas variantes, si bien se denota una tendencia en dirección de una creciente autonomía, como muestran, en diferente forma, los casos de Argentina, Brasil y Uruguay.

De Landsheere, citando a Guthrie, considera que lo ideal es que esas actividades las haga un organismo no directamente dependiente del Ministerio (apolítico), no vinculado con grupos de presión, institucionalmente creíble, al tanto de la problemática de la educación, metodológicamente competente y que tenga capacidad de sintetizar los resultados de su trabajo y de difundirlos entre los diversos grupos interesados: investigadores, especialistas en evaluación, maestros, funcionarios, padres de familia, medios de comunicación y público en general. (1994: 43)

Los resultados de las pruebas podrán incorporarse al sistema de indicadores, aportando aspectos irremplazables para el apartado de resultados, el cual es crucial para que el sistema pueda atender todas las dimensiones de la calidad educativa.



3. El subsistema de evaluación de escuelas

3.1. La escuela y la calidad

Durante mucho tiempo la educación despertó, entre la población, sentimientos curiosamente encontrados, pero igualmente carentes de fundamento sólido: por una parte, un ingenuo optimismo que hacía que se esperara de la escuela toda suerte de beneficios, de manera casi mágica: progreso económico, conciencia y participación cívica, cultura, etcétera. Los textos que fundamentan los artículos constitucionales o las leyes reglamentarias de la educación en cualquier país son representativos de este espíritu, que impregnó por igual a los franceses de la Revolución o la Comuna, a los americanos de la Independencia y la Guerra de Secesión, o a los mexicanos de la Reforma y la Revolución. Al mismo tiempo, las quejas más acerbadas sobre la baja calidad de las escuelas no son privativas de la última década del siglo XX y su fiebre neoliberal. Una tendencia a juzgar críticamente el nivel educativo de las generaciones jóvenes, en comparación con un pasado que se idealiza, se da también en México. Los maestros universitarios solemos lamentarnos con regularidad de que los alumnos “llegan cada vez peor”.

En los Estados Unidos, los inicios de la década de los años sesenta conocieron una serie de iniciativas que tenían en común el buscar una sociedad más equitativa. Fue la época de la lucha contra la segregación racial y a favor de la igualdad de derechos civiles de los negros que encabezó Martín Luther King. Después de la muerte de John F. Kennedy, el Presidente Johnson lanzó, en 1964, su iniciativa de Guerra contra la Pobreza, en el marco de la cual solicitó del Congreso la aprobación de un fuerte paquete de recursos destinados a mejorar las condiciones de infraestructura de las escuelas de alumnado mayoritariamente de color, en un

esfuerzo compensatorio que se consideraba claramente justificado.

El Congreso decidió auspiciar un importante estudio para tener más elementos de juicio al respecto, y de allí salió el Informe Coleman, cuyos resultados fueron dados a conocer en 1966. Como se esperaba, el Informe mostró claras diferencias en el nivel promedio de aprendizaje de los alumnos negros en relación con el de los blancos. Sin embargo, contra lo que se creía, no encontró diferencias significativas en cuanto a la infraestructura y el equipamiento de las escuelas con alumnado mayoritariamente de gente de color.

A resultados del estudio y, sin duda, como un elemento más del polarizado clima ideológico de fines de los sesenta (con la Guerra de Vietnam, los movimientos estudiantiles de 1965 en adelante, de Berkeley a París, etcétera) las opiniones de los investigadores educativos se polarizaron también.

Para explicar los sorprendentes resultados del Informe Coleman, algunos estudiosos subrayaban la importancia de los factores genéticos, o de los que ellos consideraban tales, sobre los resultados escolares, en tanto otros destacaban el peso de los factores asociados con la clase social de los alumnos y el medio en donde se desenvolvían. Ambas posturas, acérrimas rivales en casi todo, coincidían en un punto preciso: en considerar que la escuela era poco relevante para entender las diferencias en el aprendizaje. La frase *la escuela no cuenta*, llegó a ser el eslogan de algunos, como los teóricos de la desescolarización, Illych y Reimers.

Después, los años setenta y ochenta vieron el desarrollo de trabajos de carácter micro y enfoque cualitativo, los cuales han mostrado con toda claridad que, más allá de una posible coincidencia en lo relativo a los insumos materiales con los que



cuenta una escuela, hay aspectos mucho más sutiles que pueden incidir con fuerza en los resultados escolares. Los trabajos sobre escuelas efectivas han mostrado la importancia de variables como la existencia de:

- Un liderazgo escolar claro, centrado en lo pedagógico, con buena comunicación con los demás actores.
- Propósitos claros, valores y, en general, una cultura institucional compartida.
- Mecanismos permanentes y participativos de planeación y de evaluación-monitoreo de la calidad.
- Clima escolar tranquilo, seguro y favorecedor del aprendizaje.
- Actividades de formación permanente de los maestros.
- El que éstos tengan altas expectativas en relación con las posibilidades de aprendizaje de sus alumnos;
- Participación de los padres de familia en las actividades escolares;
- Apoyo por parte de las instancias educativas superiores (véase Thélot, 1993: 91; Hopkins, 1994: 147; cfr. también Austin y Garber, 1985; Madaus et al. 1980, etcétera).

Algunos trabajos mexicanos, como el realizado, en Puebla, por Silvia Schmelkes, han mostrado la existencia de escuelas con recursos precarios y entorno desfavorable (por ejemplo planteles bidocentes en medio rural indígena) que consiguen niveles de aprendizaje comparables a los de escuelas con mejores recursos y ambiente teóricamente más favorable, lo que pone en evidencia el peso específico de los factores llamados de la oferta educativa, en contraposición con los de la demanda (1994).

Parece claro que el aprendizaje de los alumnos es producto de la interacción compleja de variables extraescolares y escolares. El peso de las primeras es innegable, sobre todo en circunstancias de carencia extrema, pero las segundas tienen su propio peso, que no se reduce al de las otras y que significa, para los educadores, un margen significativo de acción, con posibilidades reales de incidir realmente en los resultados, más allá de los condicionamientos del entorno, si no son excesivamente desfavorables.

3.2. La evaluación de escuelas y los indicadores de la OCDE

La escuela es la unidad mejor definida para la puesta en operación de cualquier política educativa y por ello parece lógico que sea también la unidad de evaluación fundamental. Esto se aplica a la información de tipo estadístico para construir indicadores de insumos y productos, y será fundamental en lo relativo a los procesos, como se vera. El excelente estudio de Kaagan y Coley sobre los sistemas de indicadores que han creado varios estados norteamericanos, al sacar lecciones de sus experiencias para quienes emprendan esa tarea escriben, entre otras cosas: *independientemente de que un estado pretenda reportar los datos escuela por escuela debe recolectarlos y analizarlos a ese nivel. La escuela es la unidad reconocida para la oferta de los servicios educativos* (1989: 30).

En lo relativo a indicadores de insumos y productos, deberán tenerse datos de cada escuela. Ahora bien, la lista de características de escuelas efectivas de la OCDE incluye, como componentes centrales, diversas variables relativas a procesos, que no resulta sencillo capturar en un indicador estadístico. Por ello los sistemas de evaluación de los establecimientos escolares *como tales*, suelen ser de tipo más cualitativo, con amplia participación de los actores involucrados. La lista de la OCDE incluye algunos indicadores de procesos, a nivel de plantel: el relativo a las decisiones que deberían tomarse en cada escuela (C27); los del tiempo dedicado a la enseñanza, total y de ciertas materias (P11 y P12); y el relativo a las prácticas de agrupamiento de alumnos (P21). La dificultad de cuantificar aspectos finos del funcionamiento de una escuela explica el reducido número de los indicadores relativos a ellos en los que está manejando la OCDE, pero ello no significa que se desconozca su importancia.

Los trabajos de la red de la OCDE que centra la atención en la evaluación de la calidad de las escuelas como tales (véase CERI-INES 1995c) muestran una clara influencia de los trabajos sobre escuelas efectivas ya mencionados. J. Scheerens, coordinador de la citada red, clasifica los procesos en dos grupos: en el primero incluye a los que se dan a nivel de todo un centro escolar en donde incluye los relativos a la existencia de una política orientada al logro; de li-

derazgo; de planeación participativa; de currículo; y de clima ordenado. El segundo grupo se refiere a los procesos que tienen lugar en cada aula, y comprenden el tiempo efectivo en la tarea; la estructuración de la enseñanza; la oportunidad efectiva de aprendizaje; las expectativas en cuanto al desempeño de los alumnos; el grado de monitoreo de los avances; y la existencia de reforzamientos (Hopkins, 1994: 150). El autor del texto clasifica los procesos en cuatro categorías, mencionando expresamente que se basa en la literatura sobre escuelas efectivas, pero tratando de sistematizar mejor los procesos. Sus cuatro categorías se refieren a políticas y prácticas, organización, cultura y estrategias. (Hopkins, 1994: 154) Finalmente, el autor propone una lista de indicadores de procesos, como sigue:

- De estrategia: planeación, actualización de maestros, evaluación de la escuela apoyo externo.
- De políticas y prácticas: énfasis en currículo y habilidades básicas, tiempo efectivo de enseñanza, oportunidad efectiva de aprendizaje, regularidad de su monitoreo, tácticas y modelos de enseñanza (estructuración, organización del aula, tareas a realizar en casa), calidad de relaciones entre alumnos y maestros.
- De organización: liderazgo, marco de referencia claro, roles y responsabilidades definidas, trabajo en equipo, comunicación y consenso, compromiso de padres y comunidad.
- De cultura: altas expectativas, clima ordenado y de apoyo, sistema de valores compartidos, autoestima de maestros, compromiso de maestros, espíritu de colaboración de maestros.
- Relaciones externas: actualización permanente de maestros, obligaciones contractuales, organización de un sistema para el desarrollo de la capacidad institucional de acción y apoyo permanente, procesos de cambio de mediano y largo plazos, condiciones *de abrigo* en cada nivel (Hopkins, 1994: 164).

En forma independiente países han desarrollado sistemas de evaluación de establecimientos escolares, similares a los que surgieron desde principios de siglo en educación superior, con un componente de autoestudio y otro de evaluación externa. En algunos casos estos sistemas están vinculados con los mecanismos oficiales de inspección o supervisión. En los

apartados siguientes se describen brevemente algunos de estos sistemas.

3.3. El sistema inglés GRIDS

La forma más usual de promover la evaluación en el nivel de cada establecimiento escolar en el Reino Unido, en el marco de la preocupación por la calidad, consistió en *listas de control (check lists)* desarrolladas por los inspectores de las Autoridades Educativas Locales (*Local Education Authorities, LEAs*, equivalentes de los equipos de supervisión de las Zonas de Educación Básica). Según un autor, 45 de estas listas han sido publicadas desde fines de los años setenta, si bien en general no contienen orientaciones en cuanto a la forma de aplicarlas y de utilizar sus resultados (Nuttall, según De Landsheere, 1994: 129). Mucho más completo y exitoso se ha mostrado el sistema llamado *Lineamientos para la Evaluación y el Desarrollo Interno de las Escuelas*, mejor conocido por las siglas del nombre en inglés, GRIDS: (*Guidelines for Review and Internal Development in Schools*).

Un equipo del *School Curriculum Development Committee, SCDC*, comenzó a desarrollar el sistema en 1984, probándolo en más de veinte zonas educativas (LEAs), de Inglaterra y el País de Gales. En la presentación del manual del sistema, publicado en mayo de 1988, el Director del SCDC comenta que resultó notable la visión del grupo, que anticipó los cambios radicales del sistema educativo del Reino Unido implantados por la Ley de Reforma Educativa de 1988 que, con la introducción simultánea del currículo nacional y de cierto tipo de competencia entre las escuelas, hacen más importante que nunca el trabajo de cada centro, para cuya evaluación y planeación el GRIDS resulta una herramienta muy valiosa (SCDC 1988: V).

Este lineamiento se basa en una idea clave: la de que no se puede evaluar ni cambiar de golpe todos los aspectos del funcionamiento de una escuela, sino se debe proceder de manera progresiva, considerando uno o unos cuantos puntos a la vez, para continuar más tarde con otros. Naturalmente, no cualquier aspecto se presta igualmente para iniciar: algunos aspectos pueden ser considerados como más urgentes, o más sencillos, o potencialmente más fecundos.



Por consiguiente, habrá que comenzar con una evaluación muy general, que servirá simplemente para seleccionar uno o pocos aspectos para trabajar en profundidad. Luego se trabajará ese o esos aspectos, hasta llegar a una evaluación profunda y a la organización de un plan de acción (desarrollo) para mejorarlo. Esto supone que previamente se han explorado alternativas, a partir de experiencias ajenas que se pueda conocer directamente o en forma indirecta, por medio de la literatura. Terminado el ciclo podrá recomenzarse con otros aspectos. Lo deseable es que la dinámica de evaluación y desarrollo llegue a verse como parte integral de la vida cotidiana de la escuela.

El Manual del GRIDS presenta este ciclo básico y desarrolla detenidamente cada parte, ofreciendo indicaciones precisas sobre la manera de implementarlas. Los pasos fundamentales se resumen en la siguiente página. Debe señalarse que, tratándose de una evaluación *del establecimiento escolar* se supone que el director tiene un papel central en la decisión de llevar a cabo el proceso o no. En cambio, no es indispensable, que el mismo director sea también la cabeza del comité de evaluación que deberá organizarse para coordinar el proceso. Este papel puede recaer sobre alguno de los maestros el cual reúna las cualidades necesarias.

LAS CINCO ETAPAS DEL GRIDS

Etapas I: Preparación

1. Familiarícese con el modelo GRIDS, y obtenga información sobre el apoyo disponible.
2. Asegúrese de entender su papel como director.
3. Consulte a los maestros y decida con ellos lo relativo a los pasos siguientes.
4. Haga un plan de trabajo para organizar las etapas siguientes.

Etapas II: evaluación inicial general

1. Planeación.
2. Encuesta de opinión entre los maestros y, si se desea, entre padres de familia y otras personas.
3. Consenso sobre prioridades para evaluaciones específicas y acciones subsecuentes e identificación de posibles necesidades de capacitación.
4. Organización de equipos para realizar cada evaluación específica.

Etapas III: Evaluación(es) específica(s)

1. Planeación de la(s) evaluación(es).
2. Estudio de las políticas y prácticas vigentes en la escuela sobre el tema de la evaluación.
3. Investigación de políticas y prácticas alternativas (usar fuentes externas).
4. Informe de los resultados y hallazgos al equipo docente.
5. Decisiones colectivas sobre acciones específicas a desarrollar y necesidades particulares de formación.

Etapas IV: Acciones de desarrollo

1. Planeación.
2. Implementación de actividades formativas relacionadas.
3. Monitoreo de avances.
4. Evaluación de la efectividad de las acciones.

Etapas V: Evaluación general del GRIDS

1. Evaluación del proceso GRIDS en conjunto.
2. Incorporación del proceso GRIDS como parte del funcionamiento habitual de la escuela.

Después de la quinta etapa puede volverse a la segunda, haciendo de nuevo una evaluación general para seleccionar otro(s) aspecto(s) particular(es) a evaluar. Si, con los resultados de la primera evaluación general, se tienen ya otros puntos a trabajar, puede pasarse directamente a la tercera etapa. Por el contrario, si un proceso experimenta dificultades importantes en su tercera etapa, podrá suspenderse y regresarse a la segunda o, eventualmente, iniciar la siguiente con otro u otros aspectos.

3.4. Los sistemas español y francés

Francia y España han visto también, en los últimos años, el desarrollo de nuevas concepciones y sistemas para la evaluación de los establecimientos escolares. Los sistemas educativos de ambos países han sido, tradicionalmente, centralizados. A diferencia de los casos inglés (hasta 1988) o norteamericano, la noción de currículo nacional era la única posible para franceses y españoles hasta hace poco. Apenas con la Ley de 1990 se abre la posibilidad de variantes curriculares para las autonomías españolas. Con todo, la necesidad de mayor dinamismo en la ges-

tión escolar en el nivel institucional, como requisito de una mejor calidad, se plantea en estos países con la misma fuerza que en otros. El alto grado de profesionalización de su magisterio hace tal cosa, a la vez, posible e inevitable. Por ello no resulta sorprendente que en ambos países se haya desarrollado una corriente semejante que concede al centro escolar individual un margen de iniciativa y de acción considerable, mucho mayor al que la tradición centralizadora haría esperable. La forma que ha tomado esta tendencia fue la de los llamados proyectos escolares de centro.

El proyecto escolar no es otra cosa que un plan de trabajo por medio del cual una escuela retoma los grandes objetivos establecidos para la educación por la ley y el currículo nacional, y los aplica de manera concreta a su propia realidad, teniendo en cuenta las circunstancias concretas de sus maestros, alumnos y familias y los demás elementos pertinentes de su entorno. Esta concreción de los planteamientos nacionales requiere de un esfuerzo de microplaneación preciso, el cual deberá comprender un diagnóstico de la situación concreta del centro escolar de que se trate; la fijación, en consecuencia, de objetivos propios; la organización de programas de trabajo para alcanzar tales metas; y, lógicamente, un sistema de evaluación para monitorear en forma permanente los logros de la escuela en relación con los objetivos.

Dada la naturaleza del trabajo escolar, lo anterior sólo es factible si se logra un consenso alrededor de los propósitos esenciales por parte de los integrantes de la comunidad escolar, en particular los maestros, que se traduzca en un compromiso con tales propósitos y en un esfuerzo colectivo como forma habitual de operación el centro. El papel del director como líder organizacional y educativo, y no sólo como titular de una función administrativa, es un componente fundamental de esa forma de funcionamiento. Una abundante literatura pone de manifiesto la fuerza del movimiento. Pueden verse en la bibliografía las referencias a las obras de De Peretti, Emin y Lecoine y Rebinguet para Francia, así como las de Barbera, Casanova, Gairin, Sabirón, Saez-Fernández y Santos Guerra para España. La influencia del GRIDS es importante, como puede verse en el sistema QUAPE. (véase López y Darder, 1985).

3.5. Un sistema norteamericano

En los Estados Unidos surgieron en las últimas décadas una serie de sistemas de evaluación de establecimientos escolares muy diferentes a los que hemos presentado en los apartados anteriores, que también resulta de interés reseñar. Se trata de mecanismos concebidos para detectar las escuelas sobresalientes, con el propósito de darles reconocimientos y estímulos tanto materiales como simbólicos y, a la vez, permitir que otros establecimientos puedan beneficiarse de su experiencia, aprendiendo de las innovaciones que se hayan desarrollado en ellas.

Prácticamente cada estado norteamericano cuenta con algún sistema de este tipo y, además, existen otros organizados por entidades privadas. Sin embargo, el sistema más prestigioso es, sin duda, el conocido con el nombre de *Programa de Escuelas de Cinta Azul* (*Blue Ribbon School Program*, BRSP) creado por el Departamento de Educación del Gobierno de los Estados Unidos en 1983. De 82-83 a 93-94 2,809 de unas 110 mil escuelas primarias y secundarias básicas de los Estados Unidos habían merecido dicho reconocimiento: un 2.5 por ciento del total, incluyendo escuelas públicas y privadas, urbanas y rurales, pequeñas y grandes, con alumnado de todos los niveles socioeconómicos y grupos étnicos. Cada año más de 500 escuelas son escogidas en el nivel estatal y pasan al nacional; de ellas unas 300 son visitadas y algo más de 200 reciben el reconocimiento.

Además de tratarse de un sistema que reconoce a las escuelas como tales, y no ciertos programas específicos, como los de enseñanza de las Matemáticas o de las Ciencias, el BRSP sobresale entre los demás por el nivel de reconocimiento público que representa (los premios son entregados en una solemne ceremonia anual en la Casa Blanca, por el Presidente de los Estados Unidos) y por la seriedad del proceso de selección de las escuelas premiadas, que incluye una primera selección a nivel de distrito escolar, otra a nivel estatal, y una última a nivel nacional, en donde hay dos momentos de revisión de documentos y una visita a cada escuela que llega a esa etapa.

La evaluación la realiza un grupo de un centenar de personalidades que incluye a maestros reconocidos, miembros de consejos escolares, investigadores y empresarios. Se forman subgrupos de tres o cuatro



evaluadores, cada uno de los cuales revisa una veintena de solicitudes. Después de la primera revisión, alrededor de la mitad de las solicitudes pasan a la siguiente etapa, que consiste en una visita de dos días a la escuela, por parte de grupos de visitadores designados. Tanto los evaluadores como los visitadores reciben entrenamiento especial para el desempeño de sus funciones. Los visitadores redactan un reporte, que es revisado por los evaluadores, quienes formulan finalmente sus recomendaciones al Secretario de Educación, quien toma formalmente la decisión final (Odgen y Germinario, 1994: 225-263). La escala que se utiliza para evaluar, en diversos aspectos, a las escuelas que se presentan como candidatas al reconocimiento de *Escuela de Calidad Cinta Azul* tiene los siguientes valores:

Ejemplar: se califican con este valor únicamente los programas realmente excelentes, definidos como aquellas escuelas que se ubiquen dentro del 5% mejor del país en ese aspecto. Las escuelas deberían poder servir de modelo a cualquiera otra.

Fuerte: se aplica a programas muy buenos, que se sitúan claramente por arriba de la media, sin llegar al nivel del rubro anterior. Una escuela calificada con este valor debe superar por lo menos al sesenta por ciento de las demás.

Adecuada: la escuela cumple con los estándares de calidad mínimos establecidos y su nivel coincide con la media. Se subraya que las normas se definen en relación a la media.

Inadecuada: la escuela tiene deficiencias marcadas y significativas y se encuentra definitivamente por debajo de la media y de las normas establecidas.

Sin Evidencias: no hay elementos para hacer juicios sobre el aspecto de que se trate.

Los rubros que deben evaluarse, basados en las investigaciones sobre escuelas efectivas, son siete, cada uno de los cuales puede contener de tres a once incisos para formar un total de 58:

A. Liderazgo

A.1 Las grandes metas y prioridades de la escuela están claramente formuladas y parecen adecuadas. Son desarrolladas con

participación de todos los integrantes significativos de la comunidad escolar; han sido adoptadas formalmente, se revisan de manera regular y se comunican de manera efectiva a profesores, alumnos, padres de familia y a la comunidad.

A.2 El director y los maestros tienen una visión clara de la escuela y sus alumnos, que se operacionaliza en términos de objetivos específicos, políticas, programas y recursos que se requieren para alcanzar las grandes metas y prioridades. El liderazgo de la escuela ha creado un sentimiento de participación en los propósitos entre profesores, alumnos, padres de familia y personas de la comunidad, para cumplir con la misión de la escuela.

A.3 El director y los maestros tienen una comprensión común de la importancia y la estructura del liderazgo educativo dentro de la escuela. Hay lineamientos escritos sobre la responsabilidad de las diferentes funciones.

B. Ambiente de enseñanza

B.1 Los maestros participan de manera sustantiva en las decisiones relacionadas con el currículo, la enseñanza, la disciplina, la evaluación de maestros, programas y otras actividades. La participación de los maestros es efectiva en la operación de la escuela.

B.2 El personal colabora en la planeación y realización de la enseñanza. Hay oportunidades permanentes de interacción significativa y se alienta el trabajo en equipo.

B.3 Los maestros son supervisados y evaluados formal e informalmente de manera regular por personal designado para ello, éste ofrece retroalimentación útil. Se supervisa el que las evaluaciones se traduzcan en mejoras.

B.4 Se hacen provisiones especiales para el apoyo y la formación de los maestros nuevos, o de reciente ingreso a la escuela. Los procedimientos de reclutamiento y selección de profesores parecen apropiados.

B.5 Se apoya y estimula el reconocimiento a los mejores maestros tanto formal como



	informalmente y así como dentro de la escuela como fuera de ella.		
B.6	Se ofrecen oportunidades variadas para ampliar o extender los roles o funciones de los profesores, mejorar su efectividad y su satisfacción en el trabajo y disminuir su rotación.		
B.7	Un número significativo de profesores participa en actividades formativas directamente relacionadas con las prioridades de la escuela, o en programas orientados a reforzar su dominio en las cinco materias básicas.		
<i>C. Currículum y enseñanza</i>			
C.1	La diversidad de necesidades de los alumnos y de las misiones de la escuela se reflejan en la organización de ésta. Los procedimientos para asignar a los alumnos a las diversas actividades o agrupamientos de enseñanza son razonables y justos. Los alumnos tienen la posibilidad de cambiar de actividad o grupo, si sus habilidades o intereses cambian.		
C.2	Los trabajos de desarrollo curricular se traducen en una oferta sólida y rica para todos los alumnos en Inglés, Matemáticas, Ciencias, Historia y Geografía y Lenguas extranjeras.		
C.3	La escuela destaca con claridad, dentro de su oferta curricular, un núcleo básico bien concebido que se orienta claramente en el sentido de sus grandes propósitos.		
C.4	La escuela ha implementado estrategias particulares para garantizar que los alumnos aprendan efectivamente a expresarse por escrito y tiene sistemas de evaluación para ello.		
C.5	La escuela tiene programas para atender las necesidades de alumnos que requieran apoyo especial. La identificación y ubicación de esos alumnos es equitativa. Se monitorea de cerca el progreso individual y hay evidencias claras de la efectividad de los programas.		
C.6	Se ofrecen oportunidades variadas de enriquecimiento o estudio avanzado para estudiantes sobresalientes por su talento		
			o motivación. Si los alumnos que participan en ellas no representan la diversidad del alumnado se ofrece una explicación convincente.
		C.7	La biblioteca es un componente integral del programa de enseñanza de la escuela y juega un papel clave en el desarrollo de las habilidades informativas de los alumnos.
		C.8	Se hacen esfuerzos de evaluación regulares y sistemáticos, que constituyen un programa para toda la escuela y se traducen en mejoras identificables.
		<i>D. Ambiente de los alumnos</i>	
		D.1	Las políticas, prácticas y servicios de la escuela aseguran que los alumnos que entran a ella, en el primer grado o más tarde, participen con éxito en el conjunto de la vida escolar.
		D.2	La escuela utiliza incentivos, programas motivacionales y/o estrategias de enseñanza particulares, para desarrollar y mantener el interés de los alumnos por el aprendizaje.
		D.3	Hay oportunidades accesibles y variadas para que los alumnos tengan relación regular con maestros, asesores y otros adultos. Existen programas de asesoría psicopedagógica y académica cuya efectividad se evalúa sistemáticamente. El número de estudiantes que aprovecha estas oportunidades es significativo y representativo del alumnado.
		D.4	Se usan procedimientos efectivos para detectar y ofrecer asesoría y apoyo a desertores potenciales y otros alumnos de bajo rendimiento o alto riesgo. Un número significativo de los alumnos detectados recibe apoyo.
		D.5	Hay una gama de actividades extracurriculares al alcance de los alumnos. Se alienta a éstos a participar y un número sustantivo, lo hace.
		D.6	Existen normas disciplinarias razonables, explícitas y bien articuladas, que impiden la violencia e impulsan a los alumnos a comportarse correctamente, sin excesiva coerción.



- D.7 Dentro de la escuela se prohíbe el uso de armas y drogas, incluyendo el alcohol y el tabaco. Hay programas y esfuerzos amplios que llegan a la comunidad en donde se ubica la escuela para desalentar tales prácticas fuera de ella.
- D.8 Los alumnos tienen papel activo e influyen en las políticas de la escuela y del aula. Se aprecia la participación estudiantil. Esta participación es representativa del alumnado.
- D.9 Los programas, políticas y prácticas de la escuela y los maestros, propician el desarrollo de un carácter maduro, de valores democráticos, juicios éticos y conducta correcta, así como la capacidad de trabajar de manera autodisciplinada y con sentido.
- D.10 La escuela utiliza de manera efectiva estrategias curriculares y otras más para preparar a los alumnos para una vida productiva y armoniosa en una sociedad plural.
- D.11 Ídem en una sociedad globalmente competitiva.

E. Apoyo de los padres y la comunidad

- E.1 La escuela presenta evidencias de diversos tipos de participación de los padres. Un número sustantivo de éstos, representativo del conjunto, participa de hecho.
- E.2 Los avances y resultados de los alumnos y la escuela se comunican regularmente a los padres, formal e informalmente. Hay mecanismos para recibir retroalimentación.
- E.3 La escuela alienta y ayuda a los padres para que en el hogar ofrezcan un ambiente que apoye el aprendizaje de sus hijos y les informa sobre otras oportunidades de aprendizaje.
- E.4 La escuela hace esfuerzos sistemáticos para apoyar las necesidades de las familias.
- E.5 La escuela ofrece evidencia concreta de colaboración significativas con otras instituciones o grupos de la comunidad para apoyar actividades y programas escolares, promover aprendizaje extraescolar y ofrecer servicio integral a niños y familias.

F. Indicadores de éxito

- F.1 La escuela utiliza pruebas estandarizadas a nivel estatal o nacional que permiten:
- a) informar sobre los resultados de los alumnos de manera fácil de interpretar;
 - b) afirmar que los resultados de los alumnos son mejores que los de escuelas con características demográficas similares;
 - c) observar mejoras claras en los resultados de los alumnos en los últimos tres años, o resultados sobresalientes de manera regular.
- F.2 La escuela ofrece evidencias convincentes del aprendizaje de sus alumnos mediante pruebas basadas en desempeño u otras formas no tradicionales de evaluación.
- F.3 La asistencia de maestros y alumnos y el número de éstos involucrados en problemas disciplinarios serios se comparan favorablemente con los de escuelas semejantes.
- F.4 En los últimos cinco años la escuela, sus maestros y/o alumnos han recibido reconocimientos significativos que son indicadores de éxito.

G. Vitalidad organizacional

- G.1 El clima de la escuela refleja su misión y ofrece una atmósfera ordenada, con sentido, que apoya el aprendizaje, el cual respeta la diversidad y está abierta al cambio.
- G.2 Existe un proceso de planeación y mejoramiento con evidencias de liderazgo y avances.
- G.3 El personal conoce hallazgos y recomendaciones de los principales estudios de reforma educativa y las Metas Nacionales; pone en práctica o prepara cambios en consecuencia.
- G.4 En los últimos cinco años la escuela ha introducido cambios efectivos y/o superado problemas u obstáculos para la excelencia, al tiempo que mantenía los aspectos positivos.
- G.5 Se entienden de manera realista los principales retos educativos que la escuela deberá enfrentar, éstos reflejan la cuidadosa

valoración de las cambiantes necesidades del alumnado.

H. Énfasis especiales: Ciencia y Matemáticas

- H.1 La escuela ofrece un currículo en el área de ciencias que es rico en contenido, se apega a los lineamientos oficiales y estimula el interés y el desempeño de todos los alumnos.
- H.2 Ídem en Matemáticas.

3.6. Los sistemas de inspección o supervisión

Si bien, la importancia que se le da en la actualidad parece no tener precedentes, la preocupación por la calidad de las escuelas es antigua. Desde fechas tempranas los sistemas educativos la enfrentaron, y lo hicieron, por lo general, mediante sistemas de inspección y supervisión.

Aunque en este caso no parece haber relación biunívoca clara entre significativo y significado, el uso de los dos términos mencionados –inspección y supervisión escolar– alude a una duplicidad básica en la función de este tipo de mecanismos por una parte: el término *inspección*, alude preferentemente a la función evaluadora e, incluso, de control administrativo del trabajo que se hace en las escuelas; el término *supervisión*, alude más bien a la función de mejoramiento de la calidad, en la forma de apoyo técnico pedagógico.

Lillis menciona que hay poca literatura sobre los sistemas de inspección y supervisión. Este autor señala que las *funciones gemelas* de control y mejoramiento de la calidad se asignan a servicios que, a veces, son independientes del Ministerio de Educación, pero que en la mayor parte de los países dependen de él. También señala que hay pocos programas de formación para la preparación de inspectores y supervisores. (Lillis, 1992:1)

Por lo que respecta a Inglaterra y Gales, en el artículo que la *International Encyclopedia of Education* dedica al tema, Bolam explica que hay, por una parte, el Inspectorado de Su Majestad, con unos 500 inspectores con funciones de nivel nacional. Por otra parte, existe la estructura de las Autoridades Educativas Locales (LEAs), que corresponde a los distritos escolares americanos, o a las Zonas de Educación Básica en Aguascalientes (México). En Inglaterra y Gales

hay un total de 104 LEAs, que agrupan a las escuelas de áreas con una población promedio de unos 400 mil habitantes, lo cual es un tamaño considerable. Las LEAs cuentan con cierto número de supervisores cuya función es más de apoyo que de evaluación, puesto que ésta última es desarrollada principalmente por los inspectores de Su Majestad. Existen varios criterios para definir el número de supervisores-asesores que debe tener una zona. En general se considera que un supervisor no puede atender más de 20 escuelas. Exclusivamente para la función de apoyo existen, además, *Centros de Profesores* que organizan actividades de formación permanente para maestros en servicio (*In Service Educational Training*, INSET) (Bolam, 1985:4429-4432).

En Francia existe, dentro del Ministerio de Educación, un importante cuerpo de inspectores, con jerarquías precisas: los Inspectores Generales tienen un papel nacional, fundamentalmente de evaluación; los inspectores regionales y departamentales, en cambio, tienen funciones predominantemente pedagógicas (Bolam, 1985).

Hay países con inspectorados semejantes, como España o Nueva Zelanda, mientras que en otros la supervisión centra la atención en el monitoreo y el apoyo de los maestros en su trabajo en el aula, como en Alemania, con el sistema de *Schulaufsichtsbeamten*, en el nivel regional (de los *Länder*) En Suecia las escuelas privadas deben pasar por un proceso de tipo acreditación cada tres o cuatro años. Los Estados Unidos también tienen buenos sistemas de inspectorado, que operan mecanismos de evaluación en el nivel de establecimiento, como el *New York State School Quality Review*.

En América Latina los sistemas de inspección tienen gran tradición. Es posible, según Lillis, que el inspectorado de Costa Rica, establecido a principios de la década de 1830, sea el más antiguo del mundo, anterior al de Su Majestad para Inglaterra y Gales, creado en 1839. Sin embargo, en los países latinoamericanos estos sistemas suelen adolecer de vicios muy arraigados de burocratización y politización que los hacen extremadamente ineficientes. Lillis hace referencia a críticas que se hacen al inspectorado venezolano, que son idénticas a las que se escuchan en México en relación con el sistema nacional (Lillis, 1992:4-5).



El potencial de los inspectorados como elementos importantes de los sistemas de control y promoción de la calidad escolar no ha pasado desapercibido a las agencias internacionales que se ocupan de estas cuestiones. El CERI de la OCDE inició, en 1993, un programa de trabajo designado con el nombre “lo que funciona en el campo de las innovaciones” (*What works in innovation*); desde 1994 el programa centra la atención en lo relativo a la evaluación de las escuelas como unidades. Se realizaron estudios, que han producido reportes (véase por ejemplo Clímaco y Araujo, 1994).

El papel de los inspectorados en este tipo de trabajos se contempla expresamente. Trabajos como los de Deelsa (1992) y Aubegny, (1995) son indicativos de esa línea de acción. El Instituto Internacional de Planeación de la Educación de la UNESCO, comenzó a organizar, a partir de 1995, talleres regionales de formación sobre el monitoreo de la calidad de la educación primaria, enfocados al estudio del funcionamiento de las escuelas como tales, en los que se considera expresamente el lugar de los servicios de inspección y supervisión (IIPE, 1995).

En estos trabajos se discuten las ventajas y desventajas de manejar juntas o separadas la función de

control de calidad y mejoramiento. Se sacan lecciones precisas de la experiencia pasada, en el sentido de que la evaluación por sí misma no es interesante, si no va unida de alguna manera al mejoramiento; en el de que debe combinarse una evaluación externa, objetiva, con un elemento *amigable* de apoyo y consejo, por parte de pares reconocidos y capacitados; y en el sentido de que, para mejorar las escuelas, parece claro que hay que apostar a la experiencia y al profesionalismo de los maestros, haciendo de cada escuela una organización efectiva centrada en el aprendizaje y capaz de mejoramiento continuo.

3.7. Conclusión

No se utilizó la información sobre experiencias relevantes para el diseño del subsistema de evaluación de escuelas desarrolladas después de 1995. Se sabe que en los últimos años el Reino Unido ha perfeccionado su notable sistema (véase lo relativo al GRIDS), y que hizo tomando elementos de los sistemas de Suecia y Finlandia, considerados como los más avanzados de Europa en este campo.



4. Conclusión general

La revisión que sintetiza este documento sobre la situación que guardan en la actualidad los sistemas de evaluación de la educación en diversos países del mundo muestra con claridad que la actividad evaluativa es un elemento presente en las políticas educativas de todos ellos.

Aunque está claro que la revisión dista mucho de ser exhaustiva, cubre un número importante de países y, sobre todo, incluye a una buena parte de los más importantes del mundo, tanto en términos cuantitativos, por las dimensiones de su población y su sistema educativo, como cualitativos, por el nivel de desarrollo de dichos sistemas. La información menos precisa disponible permite añadir que la tendencia señalada abarca también a muchos de los países no incluidos expresamente en esta revisión, de manera que se considera hay bases suficientes para sostener la aseveración de que la evaluación es un rasgo omnipresente en las políticas educativas de la gran mayoría de los países del mundo.

A diferencia de algunas afirmaciones en sentido contrario, como se señaló en el texto, lejos de estar abandonando ciertas prácticas de evaluación, como las que utilizan pruebas de aprendizaje de aplicación masiva, dicha práctica sigue extendiéndose tanto en el país en que está más arraigada, los Estados Unidos, como en muchos otros del mundo desarrollado y de los países llamados en vías de desarrollo. También puede apreciarse que las prácticas evaluativas en modo alguno se limitan al uso de tales instrumentos, sino que utilizan diversos acercamientos. En algunos casos las pruebas se utilizan para evaluar sistemas y subsistemas educativos con propósitos de monitoreo, planeación y rendimiento de cuentas (*accountability*), y en otros para evaluar a los alumnos en lo individual, apoyando a otros mecanismos basados principalmente en la responsabi-

lidad de los maestros y manejados en cada plantel o distrito escolar.

Las críticas que suelen hacerse contra la evaluación en general, y contra las pruebas de tipo objetivo en particular, no carecen de fundamento, ya que no es raro que se manejen en forma deficiente, dando lugar a interpretaciones inadecuadas de los resultados obtenidos, que pueden tener serias consecuencias para las escuelas y los alumnos. En un nivel más profundo, es importante entender que la complejidad de los procesos educativos no puede ser captada en forma adecuada por un solo acercamiento, por lo que es necesario emplear varios, y combinarlos prudentemente para obtener una visión lo más completa posible, la cual ofrezca fundamento sólido para las decisiones y acciones de mejoramiento.

Pero concluir, a partir de las deficiencias reales o supuestas de las pruebas, que no deben de utilizarse para la evaluación de sistemas de educación básica tan grandes como el de México –el séptimo del mundo, con más de veinte millones de alumnos– es, sin duda, un error grave, que contradice la experiencia internacional más sólida. No es aceptable la falsa disyuntiva que contrapone pruebas en gran escala pobremente diseñadas a evaluaciones finas en el nivel de aula o escuela, como si no pudiera haber otras opciones. Es posible, en cambio, contar con pruebas en gran escala de excelente calidad sin duda. Su diseño y aplicación requieren de un esfuerzo técnico considerable, pero su utilización en un país como México será, seguramente, un elemento de gran valor para el mejoramiento del sistema educativo.

Por otra parte, la revisión muestra que el estatus jurídico y la ubicación organizacional de la unidad responsable de la evaluación puede ser muy variada. Como puede apreciarse en el cuadro que se presenta en la página 128, lo más usual es que sean oficinas



situadas en el seno de los ministerios quienes se encarguen de la evaluación, si bien aparecen dos tendencias:

- Por una parte, surgen entidades públicas que, si bien siguen siendo parte del ministerio, y siguen teniendo cierta dependencia con respecto a las autoridades educativas, constituyen organismos separados (descentralizados o de estatus similar) que tienen considerable independencia técnica. Es el caso del INCE español, del CITO holandés, de la ACER australiana y del INEP brasileño. Dada la peculiar situación de la autoridad educativa en el Uruguay (agencia pública que no depende del titular del ejecutivo) el estatus de la oficina encargada de la evaluación, es también de mayor independencia en relación a la que tiene normalmente una dependencia del ministerio.
- Por otra parte, algunas tareas especializadas, en especial las de mayor complejidad técnica, como la elaboración de instrumentos de evaluación de alta calidad, se encomiendan a entidades diferentes, completamente independientes del ministerio, como el ETS americano, empresa no lucrativa que maneja el *National Assessment of Education Program* de los Estados Unidos, o la agencia elaboradora de pruebas de la Universidad de Cambridge, que apoya al Ministerio de Educación de Singapur en este sentido.

Se aprecia también un alto grado de coincidencia en varios aspectos técnicos: con la excepción de Brasil y Colombia, en casi todos los países la evaluación de la educación superior se maneja en organismos diferentes a la de la educación básica, y aún en esos lugares hay considerable independencia entre ambos tipos de evaluaciones, y una extendida opinión en el sentido de que parece más conveniente separarlas.

En todos los casos se evalúan sólo algunos grados del sistema educativo, que se consideran claves, que son los últimos de niveles o ciclos. El número de grados evaluados varía desde dos hasta cinco, pero en ningún caso se evalúan mediante estos sistemas todos.

También hay coincidencia en evaluar las áreas consideradas claves del currículo, en particular las de Lengua, Matemáticas y Ciencias, si bien hay diferencias en cuanto al contenido de la última (no siempre se incluyen las mismas Ciencias Sociales). En diversos casos se estudian otras áreas, y en algu-

nos sistemas pequeños y/o muy fuertes se evalúan todas, como en Singapur, Holanda y, de otra manera, en Francia. En estos casos también la evaluación es anual y censal, puesto que tiene carácter sumativo, pero sólo en referencia a pocos grados claves: el último de la primaria, la secundaria y la enseñanza media. En la mayoría de los sistemas las evaluaciones se aplican a muestras y en ciclos multianuales, lo que es congruente con su naturaleza formativa o diagnóstica. Destaca el caso de los Estados Unidos, en donde hasta ahora el NAEP evalúa muestras nacionales de alumnos en grados y materias claves, en ciclos multianuales; a partir del año 2005, según la nueva Ley Bush, el NAEP seguirá haciendo evaluaciones muestrales, pero los estados deberán participar obligatoriamente en ellas, y además deberán hacer sus propias evaluaciones censales y anuales en los grados 4, 5, 6, 7 y 8 de la educación básica. En todos los casos los sistemas de evaluación comenzaron a funcionar con menos áreas y grados, y en forma muestral y multianual; paulatinamente se añaden áreas curriculares y grados, y en algunos casos se tiende a la evaluación censal y anual de grados claves.

Es fácil apreciar que los rasgos básicos del organismo de evaluación, los cuales se pretenden establecer en México coinciden con los que la experiencia internacional muestra como más adecuados para este tipo de entidades.

Volviendo al punto de la independencia del organismo, puede afirmarse que lo esencial no es el estatus jurídico que se tenga, pero sí su capacidad técnica y solvencia moral, de las que dependerá su credibilidad.

De Landsheere, citando a Guthrie, considera como lo ideal es que las actividades evaluativas sean desarrolladas por un organismo no directamente dependiente del Ministerio (apolítico), no vinculado con grupos de presión, institucionalmente creíble, al tanto de la problemática educativa, metodológicamente competente y que tenga capacidad de sintetizar los resultados de su trabajo y de difundirlos entre los diversos grupos interesados: investigadores, especialistas en evaluación, maestros, funcionarios, padres de familia, medios de comunicación y público en general (1994: 43).

Después de comparar los casos de España, Francia, Inglaterra, Suecia y Portugal, otro especialista

elevado nivel técnico-pedagógico, para garantizar la calidad de los procesos y la validez, fiabilidad y credibilidad de sus resultados. (Mateo, 2000: 244)

El cúmulo de experiencias revisado anteriormente, contiene sin duda, referentes suficientes para sustentar el diseño de un organismo de evaluación para México el cual tenga las características necesarias para asegurar la calidad de su trabajo y, por tanto, maximizar las posibilidades con un efecto positivo en la educación nacional.

RASGOS ESENCIALES DE ALGUNOS ORGANISMOS DE EVALUACIÓN

País	Órgano evaluador	Estatus jurídico	COBERTURA				
			Niveles educativos	Áreas curriculares	Gra-dos	Población	Periodicidad
Australia	ACER	Empresa	Básica	L-M-CN-CS	Clave	Muestra	
España	INCE	OPD	Básica	L-M-CN-CS	Clave	Muestra	Multianual
Francia	DPDP y M des Ex.	Ministerio	Básica-media	Todas	Clave	M y C	Anual
Holanda	CITO	OPD	Básica	Todas	Clave	M y C	Multianual
Singapur	Ministerio	Minist-Univ	Básica	Todas	Clave	Censo	Anual
USA	NAEP-ETS	Empresa	Básica	L-M-CN-CS	Clave	M (2005 C)	Multia. (2005 a.)
Argentina	Ministerio	Ministerio	Básica	L-M-CN-CS	Clave	M y C	Anual
Bolivia	SIMECAL	Ministerio	Básica	L-M	Clave	Muestra	Anual?
Brasil	INEP	OPD	Todos	L-M-CN-CS	Clave	Muestra	Multianual
Chile	SINCE	Minist-Univ	Básica	L-M-CN-CS	Clave	M y C	Anual
Colombia	ICFES	OPD	Todos	L-M-CN-CS	Clave	Muestra	Anual
Costa Rica	IIMEP-MEP	Minist-Univ	Básica	L-M-CN-CS	Clave	M y C	Anual
Ecuador	APRENDO	Ministerio	Básica	L-M	Clave	Muestra	Anual
Honduras	UPN	Minist-Univ	Básica	L-M	Clave	Muestra	Anual?
Perú	Ministerio	Ministerio	Básica	L-M-CN-CS	Clave	Muestra	Anual
Uruguay	UMRE	OPD	Básica	L-M-CN-CS	Clave	Censo	Anual

Abreviaturas: OPD = Organismo público descentralizado, no autónomo.
 Áreas: L = Lengua; M = Matemáticas; CN = Ciencias naturales; CS = Ciencias Sociales
 Grados: Clave=grados finales de niveles (primaria, secundaria, media).
 Población: M = Muestra; C = Censo.

Bibliografía

- Abbott R., S. Steadman y M. Birchenough (1988). *GRIDS Primary School Handbook*. Nueva York: Longman.
- AERA-APA-NCME (1999). *Standards for educational and psychological testing*. Washington, American Educational Research Association.
- Angulo Rasco J. F. (1994). Anotaciones al INCE. *Cuadernos de Pedagogía*. N° 219:33-39.
- Aubegny J. (1995). *The role of evaluation and accountability as a management tool and a means of improving effectiveness at school and local authority levels*, París: OCDE, 1995.
- Austin y H. Garber (1985). *Research on exemplary schools*. Orlando. Academic Press.
- Baker E.L. y B.H. Choppin. (1990). *Minimum Competency Testing*. En Walberg-haertel 1990: 499-502.
- Banta, Trudy W. (1991). Education Counts. *Assessment Update*. Vol.3. N° 5:3-4.
- Barbera A. V. (1990). *Método para evaluación de centros*. Madrid: Escuela Española.
- Beare H., B.J. Caldwell y R.H. Milligan (1989). *Creating an Excellent School. Some New Management Techniques*. London: Routledge.
- Boes N. Sh. (1979). Un compendio de puntos importantes dentro del movimiento de competencia mínima. En Latapi, 1991: 196-202.
- Bolam R. (1985). School Inspectors and Advisors. En Husen-Postlethwaite, 1985: 4429-4432.
- Bottani N. Y Delfau I. (1990). The Search for International Educational Indicators. *The OECD Observer*. N° 162: 14-18.
- Bracey Gerald W. (1978). Algunas dudas sobre las pruebas de competencia mínima. En Latapi, 1991: 166-175.
- Brigham, Carl C. et al. (1926). The Scholastic Aptitude Test of the College Entrance Examination Board. En Fiske, Thomas S. Ed. *The Work of the College Entrance Examination Board 1901-1925*. New York: Ginn & Co. Pp. 44-63. Citado en Donlon (1984: 2).
- Brikell Henry M. (1978) Siete notas clave sobre la evaluación de la competencia mínima. En Latapi: 1991: 135-146.
- Bryk Anthony Y K.L. Hermanson (1992). Educational Indicator System: Observations on Structure, Interpretation and Use. *Review of Research in Education*. Vol. 19: 451-484.
- Burstein et al. (1992) Education Indicators. En Alkin M.C. (Ed.) *Encyclopedia of Educational Research*. 6th. ed. Nueva York: Macmillan. Vol. 2:410-418.
- Carron Gabriel Y Ta Gnoc Chau (1992). *Implicaciones del estudio internacional sobre mejoramiento de los servicios de educación primaria para la planificación y el monitoreo*. París: IIPE.
- Casanova R. Ma. A. et al. (1991). *La Evaluación del Centro Educativo*. Curso de Formación para Equipos Directivos. Unidad Temática 5. Madrid: MEC.
- Cawelti G. (1978). *Prueba nacional de competencia. Una solución falsa*. En Latapi, P. (Coord.) 1991:147-155.
- Ceneval (1999). *Informe de resultados 1998*. México, Ceneval.
- CERI-INES (2000). *Education at a Glance. OECD Indicators*. 2000 Edition. París. OECD.
- CERI-INES (1991). *Assembly on International Education Indicators Handbook*. París: OCDE.
- CERI-INES (1995a). *Education at a Glance. OECD Indicators*. París: OECD.
- CERI-INES (1995b). *Measuring the Quality of Schools*. París: OCDE.

- CERI-INES (1995c). *Measuring What Students Learn*. París: OCDE.
- CERI-INES (1995d). *Public expectations of the final stage of compulsory ed.* París: OECD.
- CERI-INES (1991). *An International Handbook of Educational Indicators*. París: OCDE.
- Coordinación de Asesores del C. Secretario de Educación Pública (1994a). *Indicadores educativos. Una comparación entre la OCDE y México*. Mimeo, 12 pp. México: SEP.
- Coordinación de asesores del C. Secretario de Educación Pública (1994b). *Reunión académica sobre los indicadores cuantitativos del Sistema Educativo Nacional*. Doc. mecanográfico, 12 pp. México: SEP.
- Coordinación General de apoyo a la planeación. (1991) *Análisis de los indicadores del desempeño académico utilizados nacional e internacionalmente*. México: SGE-ANUIES. Mimeo. 19p.
- Cronbach, L. J., G. C. Gleser, H. Nanda Y N. Rajaratnam (1972). *The Dependability of Behavioral Measurements: Theory of Generalizability for Scores and Profiles*. Nueva York: Wiley.
- David R.G. (1990). Educational System Assessment And Planning Models. En Walberg-Haertel, 1990: 710-714.
- DEELSA (1992). *The role of the inspectorate & inspectors in the development and monitoring of school management & effectiveness*. Monograph. N° 2. París: OCDE.
- De Moura Castro, Claudio (1980). La educación en América Latina: una síntesis de los estudios comparativos sobre costos y eficiencia. *Revista Latinoamericana de Estudios Educativos*. Vol. X, N° 4, p. 121.
- Dhingra R. (1991). *Improving the information system for planning the quality of primary education: the case of India*. Monograph. N° 5. París, IIPE.
- Dirección General de Enseñanza Superior e Investigación Científica (1959). *Enseñanza Superior e Investigación Científica. Seis años de labor 1952-1958*. México: SEP.
- Dirección General de Evaluación (1997a). *Sistema Nacional de Evaluación Educativa. Características generales y propuesta de acciones inmediatas*. Mimeo. México: SEP.
- Dirección General de Evaluación (1997b). *Reseña histórica de la evaluación educativa que ha desarrollado la SEP 1970-1994*. Mimeo, Anexo DGE, 1997a. México: SEP.
- Dirección General de Evaluación (1996a). *Sistema Nacional de Evaluación Educativa. Propuesta general*. Mimeo. México: SEP.
- Dirección General de Evaluación (1996b). *Evaluación de la educación primaria. Propuesta*. Mimeo. México: SEP.
- Dirección General de Evaluación (1996c). *PARE 2. Sistematización de la evaluación educativa para elevar la calidad de la educación básica en los estados*. México: SEP.
- Donlon, Thomas F. Ed. (1984). *The College Board Technical Handbook for the Scholastic Aptitude Test and Achievement Tests*. New York: College Entrance Examination Board. Especialmente los capítulos I (The Admissions Testing Program: A Historical Overview, pp. 1-11) y II (Psychometric Methods Used in Admissions Testing Program, pp. 13-36).
- Ebel Robert L. (1978) La argumentación a favor de la prueba de competencia mínima. En Latapi, P. (Coord.) 1990: 156-165.
- (1991). *Education Counts: an Indicator System to Monitor the Nation's Educational Health*. Washington, Government Printing Office.
- Emin J.C. (1995). La mise en place d'un dispositif d'indicateurs pour le pilotage des établissements secondaires français. En CERI-INES 1995b, Cap. 10 pp. 201-231.
- Frederiksen, Norman, Robert J. Mislevy E Isaac I. Bejar Eds. (1993). *Test Theory for a New Generation of Tests*. Hillsdale, Nueva Jersey, Lawrence Erlbaum.
- Gairin Sallan J. (1993). La autoevaluación institucional como vía para mejorar los centros educativos. *Bordon*. Vol. 45. N° 3.
- Gairin Sallan J. (1995). La evaluación del proyecto educativo de centro. *Organización y gestión educativas*. N° 1:31-36.
- Grisay A. Y L. Mahlack (1991). *The quality of education in developing countries: a review of some research studies and policy documents*. París, IIPE.

- Goldstein, Harvey (1996). Introduction. *Assessment in Education: principles, policy & practice*. Vol. 3 (Julio) No. 2, pp. 125-128.
- Hallak J. (1992). *Managing schools for educational quality and equity: finding the proper mix to make it work*. Contributions N° 8. París: IIPE.
- Hopkins D. (1994). Process Indicators for School Improvement. En Tuijnman-Bottani, Cap. 8: 145-170.
- Hunt Ogden E. Y V. Germinario (1994). *The Nation's Best Schools. Blueprints for Excellence*. 2 vols. Lancaster. Technomic Publ.
- Husén, Torsten Y T. Neville Postlethwaite (1996). A Brief History of the International Association for the Evaluation of Educational Achievement (IEA). *Assessment in Education: principles, policy & practice*. Vol. 3 (July) No. 2, pp. 129-141.
- INCE (2001). *Evaluación de la Educación Secundaria Obligatoria 2000. Datos básicos*. Madrid: Ministerio de Educación, Cultura y Deporte.
- INCE (2000). *Evaluación de la Educación Primaria. Datos básicos 1999*. Madrid: Ministerio de Educación, Cultura y Deporte.
- INCE (2000). *Sistema estatal de indicadores de la educación. Síntesis 2000*. Madrid: Ministerio de Educación, Cultura y Deporte.
- INCE (1995). *Documentación básica*. MEC-SEE.
- INCE (1990). *INES NEWS. International Indicators and Evaluation of Ed. Systems*. N° 4: 1-7.
- INEP (2001). *Relatório de atividades 2000*. Brasília: Ministerio da Educação.
- JENCKS CH. et al. (1972). *Inequality. A reassessment of the effect of family and schooling in America*. Nueva York: Basic Books.
- Johnstone J. N. (1981). *Indicators of Education Systems*. Londres: Kogan Page-UNESCO.
- Kaagan S.S. Y R.J. Loby (1989). *State Education Indicators: Measured Strides, Missing Steps*. New Brunswick: Rutgers Univ. Center for Policy Research in Education.
- Keeves, J. P. (1997). Section III. Measurement in Educational Research. Introduction: Advances in Measurement in Education. En Keeves: John P. Ed. *Educational Research, Methodology, and Measurement. An International Handbook*. Oxford-New York-Tokyo, Pergamon, pp. 705-712.
- Keeves John P. (1994). *National examinations: design procedures and reporting*. París: UNESCO-IIPE. Colección Fundamentals of Educational Planning No. 40.
- Klein, Stephen P. y Laura Hamilton (1999). *Large-Scale Testing. Current Practices and New Directions*. Santa Monica, Rand Education.
- (s/f). *Laboratorio Latinoamericano de Evaluación Educativa*. Mimeografiado.
- Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (2000). *Primer estudio internacional comparativo sobre Lenguaje, Matemáticas y factores asociados, para alumnos del tercer y cuarto grado de la educación básica. Segundo informe*. Santiago de Chile: OREALC.
- De Landsheere, Gilbert (1986). *La recherche en éducation dans le monde*. París. Presses Universitaires de France. Traducida al español México: FCE, 1996.
- De Landsheere G. (1994). *Le pilotage des systèmes d'éducation*. Bruselas: De Boeck.
- Latapi, Pablo Coord. (1991). *Educación y escuela. Lecturas básicas para investigadores de la educación. II Aprendizaje y rendimiento*. México: SEP-Nueva Imagen.
- Lecointe M. Y M. Rebinguet (1990). *L'audit de l'établissement scolaire*. París: Les éditions d'organisation.
- (1994) *L'Etat de l'école*. París: Ministère de l'Éducation National.
- Lillis K.M. (1992). *Improving basic education: preconditions for successful inspection and supervision*. Increasing and improving the quality of basic ed. Monograph 7. París: IIPE.
- Livingstone I. D. (1990). Monitoring National Standards. En Walberg y Haertel, pp. 435-440.
- Lopez J.A. y P.Darder (1985). *Quafe-80. Cuestionario para el análisis del funcionamiento de la escuela*. Barcelona: Onda.
- López, Matías et al. (1938). *Medida de la inteligencia. Prueba individual de ejecución de Kobs*. México, SEP.
- Madaus G., P. Airasian Y T. Kellaghan (1980). *School effectiveness: a reassessment of the evidence*. Nueva York: McGraw-Hill.

- Maguire Th. O. (1993). Some Thoughts on the Alberta Version of Educational Quality Indicators. *The Alberta Journal of Educational Research*. Vol. XXXIX No. 2: 275-277.
- Martín, Michael O. et al Eds. (1999). *Technical Standards for IEA Studies*. Amsterdam: International Association for the Evaluation of Educational Achievement.
- Martínez Arias, Rosario (1995). *Psicometría: teoría de los tests psicológicos y educativos*. Madrid: Síntesis.
- Martínez Rizo, Felipe (1996). *La calidad de la educación en Aguascalientes. Diseño de un sistema de monitoreo*. Aguascalientes: UAA-IEA.
- Martínez Rizo, Felipe et al. (1993). *La calidad de la educación en Aguascalientes*. Aguascalientes, IEA.
- Martínez Rizo, Felipe Y Ma. Elena Escalera J. (1991). *La calidad de la educación en Aguascalientes, 1983*. México: CONALTE.
- Martínez Rodríguez, Félix Francisco (2000). El Sistema Nacional de Evaluación Educativa de México. *Revista de Educación*, No. 321, pp. 35-40.
- Mateo Andrés, Joan (2000). *La evaluación educativa, su práctica y otras metáforas*. Barcelona: Horsori-ICE U. de Barcelona.
- Mateo Andrés, Joan (1999). *Enciclopedia General de la Educación*. Barcelona, Océano. Vol. 2, Sección VI, Evaluación e investigación, pp. 529-676.
- Mcclung M.S.(1978). ¿Son justos los programas de pruebas de competencia? ¿Son legales? En Latapi, 1991: 176-187.
- McEwen N. (1993). Educational Quality Indicators. *The Alberta Journal of Educational Research*. Vol. XXXIX. No. 2:167-177.
- McMEEKIN, R. W. (1998). *Estadísticas educativas en América Latina y el Caribe*. Informe Técnico. Washington: Banco Interamericano de Desarrollo.
- Mecklenburger J. (1978). Exámenes de competencia mínima. Otra vez la moneda falsa. En Latapi, 1991: 188-195.
- Ministere De L'education Nationale (2000). *L'état de l'école*. N° 10. París: MEN.
- Ministere De L'education Nationale (2000). *Reperes, références & statistiques sur les enseignements, la formation et la recherche*. París: MEN-MR.
- Ministere De L'education Nationale (1999). *Géographie de l'école*. París: MEN.
- Ministere De L'education Nationale (1994). *L'état de l'école*. N° 4. París: MEN.
- Ministerio De Educacion (1991). *Programa de mejoramiento de la calidad en escuelas básicas de sectores pobres*. Santiago de Chile. Ministerio de Educación.
- (1991). *SIMCE 1991. Manual de interpretación de resultados*. Santiago de Chile, ME.
- Mislevy, Robert J. (1993). Foundations of a New Test Theory. En Frederiksen et al. 1993, 19-39.
- Muñiz, José (1997). *Introducción a la teoría de respuesta a los ítems*. Madrid: Pirámide.
- Muñoz Izquierdo, Carlos (1976). Costos y factores determinantes de la educación formal (plan de análisis). *Revista del Centro de Estudios Educativos*. Vol. VI, N° 2, p. 95.
- Muñoz Izquierdo, Carlos et al. (1976). Efectos de la Educación en el Sector Moderno de la Economía Urbana. (Un Estudio Piloto en la Industria Manufacturera de la Cd. de México) . *Revista del Centro de Estudios Educativos*. Vol. VI, N° 1, pp. 109-139.
- National Center For Education Statistics. (Annual) *Condition of Education. The Digest of Educational Statistics*. Washington.
- National Education Goals Panel. (Annual) *The National Education Goals Report*. Washington.
- Olson M. (1975). Measurement and Efficiency in Education. En *Educational indicators: monitoring the state of education*. Princeton. ETS.
- OREALC (2000). *Proyecto Regional de Indicadores Educativos. II Cumbre de las Américas*. Santiago de Chile. Ministerio de Educación-OREALC.
- Palafox, Juan Carlos, Juan Prawda y Eduardo Velez (1994). Primary School Quality in Mexico. *Comparativer Education Review*. Vol. 38 (Mayo) N° 2.
- Phelps, Richard P. (2000). Trends in Large-Scale Testing Outside the United States. *Educational Measurement: Issues and Practice* (Primavera), pp. 11-21.

- Pipho C. (1978). Exámenes de competencia mínima. Una mirada a los estándares del estado. En Latapi, 1991: 207-214.
- (1979). La conferencia de la NAEP sobre las pruebas de competencia mínima. En Latapi, 1991: 203-206.
- PISA (2001). *Knowledge and Skills for Life. First Results from PISA 2000*. París: OCDE.
- PISA (2000). *Measuring Student Knowledge and Skills. The PISA 2000 Assessment of Reading, Mathematical and Scientific Literacy*. París: OCDE.
- Postlethwaite T.N. (1985). International Association for the Evaluation of Educational Achievement (IEA). En Husen-Postlethwaite, 1985: 2645-2646.
- Ross K.N. (1994) *The Establishment of a Southern Africa Consortium for the Monitoring of the Quality of Education*. París: IIEP.
- Ross K. N. Y T. N. Postlethwaite (1992). *Indicators of the quality of education: a summary of a national study of primary schools in Zimbabwe*. París: UNESCO-IIPE.
- Rumberger R.W. (1994). Labour Market Outcomes as Indicators of Educational Performance. En Tuijnman-Bottani, 1994: 265-286.
- Sabiron S. F. (1990). *Evaluación de Centros Docentes*. Zaragoza: Central de Ediciones.
- SACMEQ (1995). *Southern Africa Consortium for Monitoring Educational Quality*. París, IIPE
- Saez M.J. Y M.Fernández (1993). La evaluación del proyecto educativo de un centro: el problema de los indicadores. *Revista de Ciencias de la Educación*. N° 153: 23-36.
- Santos Guerra M.A. (1993). *Estrategias para la evaluación interna de los Centros Educativos*. Madrid: MEC.
- Sauvageot C. Ed. (1992). *Primary Education in Lesotho. Indicators 1992*. París, IIPE.
- Ed. (1993). *L'enseignement fondamental au Mali. Indicateurs 1993*. París: IIPE.
- Secretaría de Educación de Guanajuato (1995). *Proyecto ejecutivo para la implementación del sistema estatal de medición de la calidad de los servicios educativos de Guanajuato*. Guanajuato: SEG.
- Schmelkes S. et al. (1992). *La calidad de la educación primaria en el Estado de Puebla. Un estudio de cinco regiones*. Mimeo.
- Special Study Panel on Education Indicators (1991). *Education Counts. An Indicator System to Monitor the Nation's Educational Health*. Washington. National Center for Educational Statistics. USA Dpt. of Education.
- Thelot C.(1993). *L'évaluation du système éducatif*. París: Nathan.
- Tiana Ferrer, Alejandro. (2000). *Cooperación internacional en evaluación de la educación en América Latina y el Caribe. Análisis de la situación y propuestas de actuación*. Washington, Banco Interamericano de Desarrollo.
- Tuijnman A. Y N. Bottani, Eds. (1994). *Making Education Count. Developing and Using International Indicators*. París: OCDE.
- Unidad de Integración Educativa (1996). *Sistema estatal de evaluación*. Monterrey: Secretaría de Educación del Estado de Nuevo León.
- Walberg H. J. (1990). National Assessment of Educational Progress: Retrospect and Prospect. En Walberg-Haertel 1990: 435-440.
- Walberg H.J. Y G.D. Haertel, Eds. (1990). *The International Encyclopedia of Educational Evaluation*. Oxford-Nueva York: Pergamon Press.
- West A.R., P.H. Wait Y S. Thomas (1994). *Community Educational Indicators Phase 2. Report to Member States*. Estrasburgo: Consejo de Europa.
- Wolff, Lawrence (1998). *Las evaluaciones educacionales en América Latina: Avance actual y futuros desafíos*. Santiago de Chile: PREAL.
- Wyatt T. (1994). Education indicators: a review of the literature. En Tuijnman y Bottani, 1994: 99-121.



ANEXO 2

BIBLIOGRAFÍA ESPECIALIZADA SOBRE EVALUACIÓN EDUCATIVA (Elaborada con apoyo de José Felipe Martínez Fernández)

1. Generalidades

- 1.1. Obras de referencia sobre investigación social y educativa.
- 1.2. Series metodológicas.
- 1.3. Fuentes bibliográficas.

2. Revistas y otras publicaciones periódicas.

- 2.1. Básicas sobre evaluación y pruebas.
- 2.2. Básicas sobre metodología y estadística.
- 2.3. Complementarias.

3. Metodología de investigación y evaluación educativa.

- 3.1. Metodología de investigación en ciencias sociales.
 - 3.1.1. En general.
 - 3.1.2. Encuestas, cuestionarios y escalas.
- 3.2. Generalidades sobre evaluación en educación.

4. Métodos Estadísticos.

- 4.1. Métodos básicos, diseños experimentales, correlación y regresión.
- 4.2. Muestreo, análisis de potencia.
- 4.3. Análisis multivariado en general, análisis factorial.
- 4.4. Modelos de ecuaciones estructurales y análisis de trayectorias.
- 4.5. Modelos jerárquicos lineales y modelos multinivel; análisis longitudinal-crecimiento.

5. Elaboración y análisis de pruebas.

- 5.1. Medición en general y en educación.
- 5.2. Desarrollo de pruebas en general; validez y confiabilidad.
- 5.3. Diseño y análisis de ítems; bancos de ítems; normalización y escalas; puntos de corte, equiparación; sesgo.
- 5.4. Teoría clásica de las pruebas y teoría de respuesta a los ítems.
- 5.5. Teoría de la Generalizabilidad.
- 5.6. Pruebas criterioles y de estándares.
- 5.7. Pruebas adaptativas, computarizadas, de ejecución y otros desarrollos.
- 5.8. Estándares de calidad de pruebas y procedimientos de metaevaluación.

6. Evaluaciones en gran escala.

- 6.1. Generalidades y metodología.
- 6.2. Experiencias particulares.



6.2.1. Sobre el College Board.

6.2.2. Sobre el NAEP.

6.2.3. Sobre la IEA.

6.2.4. Otras experiencias.

7. Indicadores.

7.1. Indicadores sociales.

7.2. Indicadores educativos en general.

7.3. Indicadores sobre educación superior.

8. Evaluación de escuelas.

I. Generalidades

I.1. Obras de referencia sobre investigación social y educativa

- Alkin, Marvin C. Ed. (1992) *Encyclopedia of Educational Research*. 6th ed. New York, Macmillan.
- Husén, Torsten Y T. S. N. Postlethwaite Eds. (1994). *The International Encyclopedia of Education*. 2d. edition. Oxford-New York, Pergamon, 12 vols.
- Husén, Torsten Y T. S. N. Postlethwaite Eds. (1991) *Enciclopedia Internacional de Educación*. 10 vols. Barcelona, Luis Vives. Traducción al español de la 1a. Edición de la obra anterior, publicada en inglés por Pergamon en 1985.
- Richardson, Virginia, Ed. (2001). *4th Handbook of Research on Teaching*. Washington, American Educational Research Association.
- Smelser, Neil J. Y Paul B. Baltes, Eds. (2001). *International Encyclopedia of the Social & Behavioral Sciences*. Oxford-New York, Pergamon, 26 vols.
- Mateo Andrés, Joan (1999). *Enciclopedia General de la Educación*. Barcelona, Océano.

I.2. Series metodológicas

Series de obras monográficas de temas metodológicos publicadas por la editorial Sage (Thousand Oaks, Ca.). Se considera conveniente adquirir todos sus números en forma retroactiva y luego mantenerlas ac-

tualizadas, por lo que se repiten en el aparatado siguiente, relativo a publicaciones periódicas. Se indica el número aproximado de títulos publicados en cada serie hasta el año 2002.

- Advanced Quantitative Techniques in the Social Sciences Series*. 10 títulos.
- Measurement Methods for the Social Sciences Series*. 10 títulos.
- Qualitative Research Methods Series*. 50 títulos.
- Quantitative Applications for the Social Sciences Series*. 140 títulos.

I.3. Fuentes bibliográficas

- Educational Testing Service. (1992). *The ETS Collection Catalog: Achievement Tests and Measurement Devices*. Vol. I. Author.
- Municio F.P. E I. Torroba. (1993) *Bibliografía sobre evaluación*, en *Revista de Ciencias de la Educación*, N° 155, pp. 151-169.
- N. B. En las referencias siguientes, los nombres de algunas editoriales se abrevian:
 ASC: Assessment Systems Corporation.
 LAE: Mahwah NJ, Lawrence Erlbaum Associates.
 Pergamon: Oxford-Nueva York, Pergamon Press.
 Sage: Beverly Hills o Thousand Oaks, Sage Publicationjs, Inc.
 Wiley: Nueva York, John Wiley & Sons.

2. Revistas y otras publicaciones periódicas

2.1. Básicas sobre evaluación y pruebas

ACT Research Reports. (Gratuita)
Applied Measurement in Education. LEA. ISSN 0895-7347.
Applied Psychological Measurement. Sage. ISSN 0146-6216.
Condition of Education. The Digest of Educational Statistics. Annual. National Center for Education Statistics, Washington.
CRESST Research Reports.
Educational Assessment. LEA. ISSN 1062-7197.
Educational Evaluation & Policy Analysis.
Educational Measurement: Issues & Practice. ISSN 0731-1745. NCME.
Educational & Psychological Measurement. Sage. ISSN 0013-1644.
Education at a Glance: OCDE Indicators.
Evaluation Review. A Journal of Applied Social Research. Sage. ISSN 0193-841X.
ETS Research Reports.
International Journal of Testing. LEA. ISSN 1530-5058.
Journal of Educational Measurement. NCME.
Journal of Educational Statistics.
Measurement. Lawrence Erlbaum. (A partir de 2003).
Mental Measurements Yearbook. Buros Institute.
National Education Goals Report. Annual. National Education Goals Panel. Washington.
Psychometrika.
Serie Investigación y Evaluación Educativa. Colombia. ICFES. Servicio Nacional de Pruebas.

2.2 Básicas sobre metodología y estadística

Advanced Quantitative Techniques in the Social Sciences Series. Sage.
Journal of Multivariate Analysis.
Measurement Methods for the Social Sciences Series. Sage.
Multivariate Behavioral Research. LEA. ISSN 0027-3171.
Qualitative Inquiry. Sage. ISSN 1077-8004.
Qualitative Research Methods Series. Sage.
Quality & Quantity: International Journal of Methodology. Kluwer.
Quantitative Applications for the Social Sciences Series. Sage.
Sociological Methodology. ASA.
Sociological Methods & Research. Sage. ISSN 0049-1241.
Structural Equation Modelling. LEA. ISSN 1070-5511.

2.3. Complementarias

British Journal of Mathematical & Statistical Psychology.
Journal of Econometrics.
Journal of the ASA.
New Directions in Testing & Measurement. Jossey Bass. Serie suspendida.
Popular Measurement. Journal of the Inst. for Objective Measurement. <http://mesa.spc.uchicago.edu>
Psychological Methods. APA. ISSN 1082-989X.
Review of Educational Research. AERA.

3. Metodología de investigación, evaluación y medición

3.1. Metodología de investigación en ciencias sociales

3.1.1. En general

- Anguera, Ma. Teresa *et al.* (1998). *Métodos de investigación en psicología*. Madrid, Síntesis.
- Bickman, Leonard (2000). *Validity & Social Experimentation*. Sage.
- Bickman, Leonard Ed. (2000). *Research Design. Donald Campbell's Legacy*. Sage.
- Bickman, Leonard y Debra J. Rog (1998). *Handbook of Applied Social Research Methods*. Sage.
- Cohen, Louis (1976). *Educational Research in Classrooms and School: A Manual of Materials and Methods*. Londres-Nueva York, Harper & Row Publicaciones.
- Delgado, Juan Manuel Y J. Gutiérrez, Coord. (1995). *Métodos y técnicas cualitativas de investigación en ciencias sociales*. Madrid, Síntesis.
- Denzin, Norman K. e Yvonna S. Lincoln. (2000). *Handbook of Qualitative Research*. 2d. ed. Sage.
- Keeves, John P. (1997). *Educational Research, Methodology and Measurement. An International Handbook*. Pergamon.
- King, Gary, Robert O. Kehoane Y Sidney Verba (2000). *El diseño de la investigación social. La inferencia científica en los estudios cualitativos*. Madrid, Alianza.
- Miller, (1995). *Handbook of Social Research and Measurement*. Sage.
- Nowakowski, Jeri R. Y M. A. Bunda (1984). *Handbook of Educational Variables. A Guide to Evaluation*. Dordrecht, Kluwer.
- Thompson, Bruce Ed. (1999). *Advances in Social Sciences Methodology*. Vol. 5. Stanford CO. JAI Press.

3.1.2. Encuestas, cuestionarios y escalas

- Aiken, Lewis R. (1998). *Tests & Examinations: Measuring Abilities and Performance*. ASC.
- (1997). *Questionnaires and Inventories: Surveying Opinions and Assessing Personality*. ASC.
- (1996). *Rating Scales and Checklists: Evaluating Behavior, Personality and Attitudes*. ASC.
- Bradburn, Norman M. Y Seymour Sudman (1988). *Polls and Surveys. Understanding What They Tell Us*. San Francisco, Jossey Bass.
- Fink, Arlene (1995). *How to Analyze Survey Data*. ASC.
- Fowler, Floyd J. (1995). *Improving Survey Questions: Design and Evaluation*. ASC.
- (1993). *Survey Research Methods*. ASC.
- Litwin, Mark S. (1995). *How to Measure Survey Reliability and Validity*. ASC.
- Mciver, John P. Y E. G. Carmines (1981). *Unidimensional Scaling*. ASC.
- Rea, Louis M. Y Richard A. Parker (1992). *Designing and Conducting Survey Research. A Comprehensive Guide*. San Francisco, Jossey Bass.
- Rossi, P. (1985). *Handbook of Survey Research*. ASC.
- Schuman, Howard Y S. Presser (1996). *Questions and Answers in Attitude Survey: Experiments on Question Form, Wording and Context*. ASC.
- Spector, Paul E. (1992). *Summated Rating Scale Construction: An Introduction*. ASC.
- Weisberg, Herbert F., J. A. Krosnick Y B. D. Bowen (1996). *An Introduction to Survey Research, Polling and Data Analysis*. ASC.
- Generalidades sobre evaluación en educación
- Finck, Frederick L. (1997). *Educational Performance Assessment*. Riverside Publicaciones Houghton Mufflin Co.



- Goldstein, Harvey (1996). Introduction. *Assessment in Education: principles, policy & practice*. Vol. 3 (Julio) No. 2, pp. 125-128.
- Madaus, G. Y D. Stufflebeam (1989). *Educational Evaluation: Classic Works of Ralph W. Tyler*. Dordrecht, Kluwer.
- Madaus G., M. Scriven Y D. Stufflebeam (1983) *Evaluation Models*. Dordrecht-Boston. Kluwer-Nijhoff.
- Murphy Roger Y H. Torrance (Eds.) (1988) *Evaluating Education: Issues and Methods*. London. Paul Chapman.
- Stufflebeam D.L. Y Shinkfield A.J. (1987) *Evaluación sistemática. Guía teórica y práctica*. Barcelona. Paidós-MEC.
- Walberg H.J. Y G.D. Haertel (Eds.) (1990) *The International Encyclopedia of Educational Evaluation*. Pergamon.
- Wilson J.D. (1988) *Cómo valorar la calidad de la enseñanza*. Barcelona, Paidós-MEC.

4. Métodos Estadísticos

- Métodos básicos, diseños experimentales, correlación y regresión
- Agresti A. (1996). *Introduction to Categorical Data Analysis*. John Wiley.
- Agresti, A. (1990). *Categorical Data Analysis*. John Wiley.
- Agresti, A. (1984). *Analysis of Ordinal Categorical Data*. John Wiley.
- Allen, Michael Patrick (1997). *Understanding Regression Analysis*. ASC.
- Andrews, Frank M. et al. (1998). *Selecting Statistical Techniques for Social Science Data : A Guide for SAS Users*. Cary NC, SAS Institute.
- Cliff, Norman (1996). *Ordinal Methods for Behavioral data Analysis*. LEA.
- Cohen, Jacob Y P. Cohen (1983). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. 2d. ed. ASC.
- Cotton, John W. (1997). *Analyzing Within-Subjects Experiments*. LEA.
- Foster, Jeremy (2000). *Data Analysis Using SPSS for Windows Version 8.0-10.0. A Beginner's Guide*. Sage.
- Fox, John (1997). *Applied Regression Analysis, Linear Models and Related Methods*. ASC.
- Freund, Rudolf F. Y W. J. Wilson (1998). *Regression Analysis: Statistical Modeling of a Response Variable*. ASC.
- Hoaglin, Mosteller Y Tukey. *Exploratory Approach to Analysis of Variance*. John Wiley.
- Hoaglin, Mosteller Y Tukey. *Exploring Data Tables, Trends and Shapes*. John Wiley.
- Hoaglin, Mosteller Y Tukey. *Understanding Robust & Exploratory Data Analysis*. John Wiley.
- Hollander, Myles Y D. A. Wolfe (1999). *Nonparametric Statistical Methods*. ASC.
- Holmer, D. W. Y S. Lemeshow (1989). *Applied Logistic Regression*. John Wiley.
- Jacoby, William G. (1997). *Statistical Graphics for univariate and Bivariate Data*. ASC.
- Jaeger, Richard M. (1990). *Statistics: A Spectator Sport*. ASC.
- Kanji, Gopal K. 100 *Statistical Tests*. Sage.
- Kirk, Roger E. (1999). *Statistics: An Introduction*. ASC.
- Le, Chap T. (1998). *Applied Categorical Data Analysis*. ASC.
- Little, R. J. Y D. B. Rubin (1987). *Statistical Analysis with Missing Data*. John Wiley.
- Long, J. Scott (1997). *Regression Models for Categorical and Limited Dependent Variables: Analysis and Interpretation*. ASC.
- Maxwell, Scott E. Y H. D. Delaney (2000). *Designing Experiments and Analyzing Data. A Model Comparison Perspective*. LEA.
- Myers, Jerome L. Y A. D. Well (1995). *Research Design and Statistical Analysis*. LEA.
- Morgan, George A., O. V. Griego Y G. W. Gloeckner (2000). *Spss For Windows. An Introduction to Use and Interpretation in Research*. LEA.
- Palmer Pol, Alfonso L. (1999). *Análisis de Datos. Etapa Exploratoria*. Madrid, Pirámide.
- Pedhazur, Elazer J. (1982). *Multiple Regression in Behavioral Research*. Nueva York, Holt, Rinehart & Winston.
- Pedhazur, Elazer J. Y L. Pedhazur-Schmelkin (1991). *Measurement, Design and Analysis: An Integrated Approach*. LEA.
- Rudas, Tamás (1997). *Odds Ratios In The Analysis Of Contingency Tables*. Asc.
- Rutherford, Andrew (2000). *Introducing Anova And Ancova. A General Linear Modeling (GLM) Approach*. Sage.
- Sirkin, R. Mark (1994). *Statistics For The Social Sciences*. Asc.

- Tukey, (1977). *Exploratory Data Analysis*. Reading, Addison Wesley.
- Vogt, W. Paul (1998). *Dictionary Of Statistics & Methodology. A Nontechnical Guide for the Social Sciences*. Sage.
- Von Eye, Alexander Y Ch. Schuster (1998). *Regression Analysis For The Social Sciences*. ASC.
- Walgreen, Anders et al. (1996). *Graphing Statistics And Data: Creating Better Charts*. ASC.
- Wilcox, Rand R. (1995). *Statistics for the Social Sciences*. ASC.

4.2 Muestreo, análisis de potencia (power analysis).

- Chmura Kraemer, Helena Y S. THIEMANN (1987). *How Many Subjects Statistical Power Analysis in Research*. ASC.
- Cochran. *Sampling Techniques*. John Wiley.
- Cohen, Jacob (1988). *Statistical Power Analysis for the Behavioral Sciences*. LEA.
- Hansen, Hurwitz y Madow. *Sample Survey Methods and Theory. Vol I: Methods and Applications. Vol. II: Theory*. John Wiley.
- Levy y Lemeshow. *Sampling of Populations: Methods and Applications*. John Wiley.
- Murphy, Kevin R. Y B. Myers (1998). *Statistical Power Analysis: A Simple and General Model for Traditional and Modern Hypothesis*. LEA.
- Thompson. *Sampling*. Wiley.
- Walner, Howard (2000). *Drawing Inferences form Self Selected Samples*. LEA.
- Westfall Y. Young. *Resampling Based Multiple Testing*. Wiley.

Análisis multivariado en general, análisis factorial (factor analysis).

- Arminger, G. C. C. Clogg Y M. E. Sobel Eds. (1995). *Handbook of Statistical modeling for the social and behavioral sciences*. New York, Plenum Press.
- Bartolomew, D. J. (1987). *Latent Variable Models And Factor Analysis*. Nueva York, Oxford University Press.
- Carroll, J. Douglas, P. E. Green Y A. Chaturvedi (1997). *Mathematical Tools for Applied Multivariate Analysis*. ASC.

- Comrey, Andrew L. Y H. B. Lee (1992). *A First Course In Factor Analysis*. Asc.
- Garson, G. David (1998). *Neural Networks: An Introductory Guide for Social Scientists*. ASC.
- Gorsuch, Richard L. (1983). *Factor Analysis*. Lea.
- Grimm, Lawrence G. Y P. R. Yarnold, Eds. (1995). *Reading and Understanding Multivariate Statistics*. Washington, APA.
- Hartman, H. H. (1976). *Modern Factor Analysis*. Chicago, The University of Chicago Press.
- Lindsey, James (1997). *Applying Generalized Linear Models*. Asc.
- Marcoulides, George A. Y S. L. Hershberger (1997). *Multivariate Statistical Methods: A First Course*. LEA.
- Stevens, James P. (2002). *Applied Multivariate Statistics For The Social Sciences*. LEA.
- Tacq, Jacques (1997). *Multivariate Analysis Techniques in Social Science Research: From Problem to Analysis*. ASC.
- Tinsley, Howard E. Y S. D. Brown, Eds. (2000). *Handbook of Applied Multivariate Statistics and Mathematical Modeling*. ASC.
- Von Eye, Alexander Y K. E. Niedermeier (1999). *Statistical Analysis Of Longitudinal Categorical Data in the Social and Behavioral Sciences*. ASC.
- Waller, Niels G. Y P. E. Meehl (1997). *Multivariate Taxometric Procedures: Distinguishing Types from Continua*. ASC.
- Wickens, Thomas D. (1994). *The Geometry of Multivariate Statistics*. ASC.

Modelos de ecuaciones estructurales (Structural Equations Modelling, SEM) y análisis de trayectorias (Path Analysis).

- Bollen, Kenneth A. Y J. S. Long Eds. (1993). *Testing Structural Equation Models*. ASC.
- Byrne, Barbara M. (2001). *Structural Equation Modeling with AMOS*. Lawrence Erlbaum A.
- Byrne, Barbara M. (1998). *SEM with LISREL, PRELIS and SIMPLIS: Basic Concepts, Applications and Programming*. ASC.
- Hoyle, Rick H. (1995). *SEM: Concepts, Issues and Applications*. ASC.
- Jöreskog, Karl G. Y Dag Sörbom (1993). *LISREL 8: SEM with the SIMPLIS Command Language*. ASC.

- Kaplan, David (2000). *Structural Equation Modeling: Foundations and Extensions*. Sage.
- Kelloway, E. Kevin (1998). *Using LISREL for SEM: A Researcher's Guide*. ASC.
- Loehlin, John C. (1992). *Latent Variable Models: An Introduction to Factor, Path, and Structural Analysis*. LEA.
- Marcoulides, George A. Y Irini Moustaki, (2002). *Latent Variable and Latent Structure Models*. Lawrence Erlbaum Assoc.
- Marcoulides, George A. Y Randall E. Schumacker (2001). *New Developments and Techniques in Structural Equation Modelling*. Lawrence Erlbaum Assoc.
- Marcoulides, George A. (2001). *Advanced Structural Equation Modeling: New Developments and Techniques*. LEA.
- Marcoulides, George A. Y Randall E. Schumaker (1996). *Advanced Structural Equation Modeling: Issues and Techniques*. LEA.
- Maruyama, Geoffrey M. (1997). *Basics of Structural Equation Modeling*. ASC.
- Raykov, Tenko Y George A. Marcoulides (2000). *A First Course in Structural Equation Modeling*. Lawrence Erlbaum Assoc.
- Schumacker, Randall E. Y George A. Marcoulides, Eds. (1998). *Interaction and Nonlinear Effects in Structural Equation Modeling*. Lawrence Erlbaum Assoc.
- Schumacker, Randall E. y Richard G. Lomax (1996). *A Beginner's Guide to Structural Equation Modeling*. Lawrence Erlbaum Assoc.
- Modelos jerárquicos lineales (Hierarchical Linear Modeling, HLM) y multinivel (Multilevel Modeling, MM); análisis longitudinal o de crecimiento (growth modeling)**
- Bryk, A. S. Y S. W. Raudenbush (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, Sage.
- Collins, Linda M. Y J. L. Horn Eds. (1991). *Best Methods for the Analysis of Change. Recent Advances, Unanswered Questions, Future Directions*. Washington, APA.
- Duan, N. Y S. Reise Eds. (2001). *Multilevel Modeling: Methodological Advances, Issues and Applications*. LEA.
- Duncan, Terry E. et al. (1999). *Introduction to Latent Variable Growth Curve Modeling*. LEA.
- Franklin, Ronald D., D. B. Allison Y B. S. Gorman (1997). *Design and Analysis of Single-Case Research*. LEA.
- Goldstein, H. (1995). *Multilevel Statistical models*. London, Edward Arnold.
- Gottman, John M. (1995). *The Analysis of Change*. ASC.
- Heck, Ronald H. Y S. L. Thomas (2000). *An Introduction to Multilevel Modeling Techniques*. ASC.
- Hox, Joop (2002). *Multilevel Analysis*. Mahwah, NJ, Lawrence Erlbaum Assoc.
- Kreft, Ita G. G. Y J. De Leeuw (1998). *Introducing Multilevel Modeling*. ASC.
- Little, Todd D., Kai U. Schnabel Y Jürgen Baumert (2000). *Modeling Longitudinal and Multilevel Data*. Lawrence Erlbaum Assoc.
- Moskowitz, Debbie S. Y Scott L. Hersher (2002). *Modeling Intraindividual Variability with Repeated Measures Data*. Lawrence Erlbaum Assoc.
- Reise, Steven P. Y Naihua Duan, Eds. (2002). *Multilevel Modeling*. Lawrence Erlbaum A.
- Sayer, A. Y L. Collins (1998). *New Methods for the Analysis of Change*. Washington, APA.
- Snijders, T. Y R. Bosker (1999). *Multilevel analysis. An introduction to basic and advanced multilevel modeling*. Sage.
- Taris, Toon W. (2000). *A Primer in Longitudinal Data Analysis*. Sage.
- Van Der Kamp, Leo (1998). *Longitudinal Data Analysis: Designs, Models & Methods*. ASC.
- Von Eye, Alexander (1990). *Statistical Methods in Longitudinal Research. Vol. I: Pples. and Structuring Change. Vol. II: Time Series and Categorical Longitudinal Data*. ASC.
- Von Eye, Alexander Y C. C. Clogg (1996). *Categorical Variables in Developmental Research: Methods of Analysis*. ASC.
- Von Eye, Alexander y K. E. Niedermeier (1999). *Statistical Analysis of Longitudinal Categorical Data in the Social and Behavioral Sciences*. LEA.

5. Elaboración y Análisis de pruebas (Testing)

5.1. Medición en general y en educación

- Allen, M. J. Y W. M. Yen (1979). *Introduction to Measurement Theory*. Monterey, Ca., Brooks/Cole.
- Anderson, Lorin Y Sid F. Bourke (2000). *Assessing Affective Characteristics in the Schools*. Lawrence Erlbaum Assoc.
- Bartholomew, David J. (1996). *Statistical Approach to Social Measurement*. ASC.
- Birnbaum, Michael H. Ed. (1997). *Measurement, Judgement and Decision Making*. ASC.
- Braun, Henry, D. N. Jackson Y D. E. Wiley (2001). *The Role of Constructs in Psychological and Educational Measurement*. LEA.
- Campbell, Donald T. Y M. J. Russo (2000). *Social Measurement*. Sage.
- Embreston, Susan E. Y Scott L. Hershberger Eds. (1999). *The New Rules of Measurement. What Every Psychologist and Educator Should Know*. LEA.
- Guion, Robert (1997). *Assessment, Measurement and Prediction for Personnel Decisions*. ASC.
- Keeves, J. P. (1997). Section III. Measurement in Educational Research. Introduction: Advances in Measurement in Education. En Keeves, John P. Ed. *Educational Research, Methodology, and Measurement. An International Handbook*. Oxford-New York-Tokyo, Pergamon, pp. 705-712.
- Krantz, D. H. et al. (1971). *Foundations of Measurement*. Vol. I, Nueva York, Academic Press.
- Laveault, Dany et al. (1994). *Modern Theories of Measurement : Problems & Issues*. ASC.
- Linn, Robert L. (1989). *Educational Measurement*. Nueva York, Macmillan.
- Luce, R. D. et al. (1990). *Foundations of Measurement*. Vol. III. New York, Academic Press.
- Mehrens, W. A. Ed. (1976). *Readings in measurement and evaluation in education and psychology*. Nueva York, Holt, Rinehart & Winston.
- Michell, J. (1990). *An introduction to the logic of psychological measurement*. L. Erlbaum.
- Pawson, R. (1989). *A Measure for Measures*. Londres, Routledge.
- Popham, W. James (1981). *Modern educational measurement*. Englewood Cliffs, Prentice-Hall.
- Robertss, F. S. (1979). *Measurement theory*. Reading, Addison Wesley.
- Schank, Roger C. (2001). *Scrooge Meets Dick and Jane*. Lawrence Erlbaum.
- Suppes, Patrick et al. (1988-1990). *Foundations of Measurement*. Vols. III. Nueva York, Academic Press. ASC.
- Thorndike, R. M., G. K. Cuningham, R. L. Thorndike Y E. P. Hagen (1991). *Measurement and evaluation in psychology and education*. Nueva York, Maxwell McMillan.
- Thorndike, R. L. (1982). *Applied Psychometrics*. Boston, Houghton-Mifflin.
- Wainer, Howard Y Samuel Messick (1983). *Principals of Modern Psychological Measurement*.
- Ward, Stocker Y M. Ward (1995). *Educational Measurement. 2 Vols*. University Press.
- Zeller, R. A. Y E. G. Carmines (1979). *Measurement in the social sciences: the link between theory and data*. New York, Cambridge University Press.

5.2. Desarrollo de pruebas en general; validez y confiabilidad

- Anastasi, A. Y S. Urbina (1997). *Psychological Testing*. 6th. Ed. Nueva York, McMillan. Trad. española Aguilar, 1973.

- Anderson, Scarvia B. Y J. S. Helmick, Eds. (1983). *On Educational Testing*. San Francisco, Jossey Bass.
- Bejar, Isaac I. (1983). *Achievement Testing: Recent Advances*. ASC.
- Bellack, A. S. Y M. Hersen Eds. (1988). *Behavioral Assessment. A Practical Handbook*. 3d. ed. Nueva York, Pergamon.
- Buela Casal, Gualberto Y J. Carlos Sierra Dirs. (1997). *Manual de evaluación psicológica. Fundamentos, técnicas y aplicaciones*. Madrid, Siglo XXI.
- Carmines, Edward G. Y R. A. Zeller (1979). *Reliability and Validity Assessment*. ASC.
- Cronbach, Lee J. (1970). *Essentials of Psychological Testing*. Nueva York, Harper & Row.
- Goldstein G. Y M. Hersen (1990). *Handbook of Psychological Assessment*. 2d. ed. Nueva York, Pergamon.
- Kline, Paul (1986). *A Handbook of Test Construction. Introduction to Psychometric Design*. Londres, Methuen.
- Martínez Arias, Rosario (1995). *Psicometría: teoría de los tests psicológicos y educativos*. Madrid. Síntesis.
- Muñiz, José Coord. (1996). *Psicometría*. Madrid, Universitas.
- Nunnally, J. C. (1967). *Psychometric Theory*. Nueva York, Mc Graw Hill.
- Popham, W. James (1975). *Educational evaluation*. Englewood Cliffs NJ, Prentice-Hall.
- Ronning, R. R. et al. (1987). *The influence of cognitive psychology on testing*. LEA.
- Thissen, David Y Howard Wainer (2001). *Test Scoring*. Lawrence Erlbaum.
- Traub, Ross E. (1994). *Reliability for the Social Sciences: Theory and Applications*. ASC.
- Wainer, Howard Y H. I. Braun (1988). *Test Validity*. LEA.
- Walsh, W. Bruce y N. E. Betz (1995). *Tests and Assessment*. ASC.
- Berk, R. A. Ed. (1982). *Handbook of Methods for detecting item bias*. Baltimore, John Hopkins.
- Camilli, Gregory y Lorrie A. Shepard (1994). *Methods for Identifying Biased Test items*. MMSS No. 4. Newbury Park, Sage.
- Haladyna, Thomas M. (1994). *Developing and Validating Multiple-Choice Test Items*. LEA.
- Holland, P. W. Y D. B. Rubin Eds. (1982). *Test equating*. Nueva York, Academic Press.
- Holland, Paul W. Y H. Wainer Eds. (1993). *Differential Item Functioning*. LEA.
- Irvine, Sidney Y Patrick Kyllonen (2002). *Item Generation for Test Development*. LEA.
- Jensen, A. R. (1980). *Bias in mental testing*. Nueva York, The Free Press.
- Kolen, Michael J. Y R. L. Brennan (1995). *Test Equating: Methods and Practices*. ASC.
- Livingstone, S. A. Y M. J. Zieky (1982) *Passing scores*. Princeton, N. J. ETS.
- Osterlind, Steven J. (1997). *Constructing Test Items*. Dordrecht, Kluwer.
- Osterlind, Steven J. (1983). *Test item bias*. Beverly Hills, Sage.
- Potenza, Maria T. Y Martha L. Stocking (1994). *Flawed Items in Computerized Adaptive Testing*. Princeton, ETS, 46 pp. ED 385 556.
- Reynolds, C. R. Y R. T. Brown Eds. (1984). *Perspectives on bias in mental testing*. Nueva York, Plenum Press.
- Roid, G. Y T. Haladyna (1981). *A Technology for Test-Item Writing*. ASC.
- Wilheingham W. W. Y Nancy S. COLE (1997). *Gender and Fair Assessment*. LEA.
- Wongbundhit, Yuwadee (1985). *Item Banking Procedure and Quality Control in Dade County Public Schools*. Annual Meeting of the AERA, 17 pp. ED 262 048.

5.3. Diseño y análisis de ítems; bancos de ítems (item banking); normalización y escalas; puntos de corte; equiparación; sesgo

- Angoff, W. H. (1984). *Scales, norms and equivalent scores*. Princeton, NJ, ETS.

5.4. Teoría clásica de las pruebas y teoría de respuesta a los ítems

- Andrich, David (1988). *Rasch Models for Measurement*. ASC.
- Baker, Frank B. (1992). *IRT: Parameter Estimation Techniques*. ASC.
- Baker, Frank B. (1985). *The basis of Item Response Theory*. Portsmouth, Heineman.

- Bond, Trevor G. Y Christine M. Fox (2001). *Applying the Rasch Model*. Mahwah, NJ, Lawrence Erlbaum Assoc.
- Crocker, L. y J. Algina (1986). *Introduction to classical and modern test theory*. Nueva York, Holt, Rinehart & Winston.
- Embreston, Susan E. Y Steven P. Reise (2000). *Item Response Theory for Psychologists*. Mahwah, NJ, Lawrence Erlbaum Assoc.
- Fischer, Gerhard H. E I. W. Molenaar (1995). *Rasch Models: Foundations, Recent Developments and Applications*. ASC.
- Frederiksen, Norman, Robert J. Mislevy e Isaac I. BEJAR Eds. (1993). *Test Theory for a New Generation of Tests*. Hillsdale, Nueva Jersey, Lawrence Erlbaum.
- Gulliksen, H. (1950). *Theory of mental tests*. Nueva York, Wiley.
- Hambleton, Ronald K., H. Swaminathan Y H. Jane Rogers (1992). *Fundamentals of Items Response Theory*. MMSS Series No. 2. Newbury Park, Sage. ASC.
- Hambleton, Ronald K. Y J. N. Zaal Eds. (1991). *Advances in Educational and Psychological Testing: Theory and Applications*. Dordrecht, Kluwer.
- Hambleton, Ronald K. Y Hariharan Swaminathan (1985). *Item Response Theory. Principles and Applications*. Boston-Dordrecht-Lancaster, Kluwer. ASC.
- López-Piña, J. A. (1995). *Teoría de respuesta al ítem: fundamentos*. Barcelona. PPU.
- Lord, Frederic M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ, Lawrence Erlbaum. ASC.
- Lord, Frederick M. Y Melvin R. Novick (1968). *Statistical Theories of Mental Test Scores*. Reading, Addison Wesley.
- Lord, Frederic M. (1951). *A Theory of Test Scores*.
- McDonald, Roderick P. (1999). *Test Theory. A Unified Treatment*. Lawrence Erlbaum.
- Mislevy, Robert J. (1993). Foundations of a New Test Theory. En Frederiksen et al. 1993, pp. 19 a 39.
- Rasch, Georg (1992). *Probabilistic Models for Some Intelligence and Attainment Tests*. ASC.
- Suen, Hoi K. (1990). *Principles of Test Theory*. LEA. ASC.
- Van Der Linden, Wim J. Y R. K. Hambleton (1997). *Handbook of Modern IRT*. ASC.
- Wilson, Mark Ed. *Objective Measurement. Theory into Practice*. Ablex Publ. Co. – JAI Press. Vol. 1, 1992; Vol. 2, 1994; Vol. 3, 1996; Vol. 4, 1997.
- Wright Benjamin D. Y Mark H. Stone (1998). *Diseño de mejores pruebas utilizando la técnica de Rasch*. México. Ceneval. (Best test design. Mesa Press, Chicago).

5.5. Teoría de la Generalizabilidad

- Brennan, R. L. (1983). *Elements of Generalizability Theory*. Iowa City, ACT.
- Brennan, R. L., D. Jarjoura y E. L. Deaton (1980). *Some issues concerning the estimation and interpretation of variance components in generalizability theory*. ACT Technical Bulletin No. 36. Iowa City, ACT.
- Cardinet, J. E y. Tourneur (1985). *Assurer la mesure. Guide pour les études de generalizabilité*. Berna, Lang.
- Cronbach, Lee J., G. Gleser, H. Nanda y N. Rajaratnam (1972). *The Dependability of Behavioral Measurements: Theory of Generalizability for Scores and Profiles*. Nueva York, Wiley. (University Microfilms)
- Shavelson, Richard J. y Noreen M. Webb (1991). *Generalizability Theory. A Primer*. MMSS Series No. 1. Newbury Park, Sage.
- Webb, Noreen M., R. J. Shavelson y E. Maddahian (1983). Multivariate generalizability theory. *New directions in Testing & Measurement*, No. 18. San Francisco, Jossey Bass.

5.6. Pruebas criterios y de estándares

- Berk, Ronald A. Ed. (1984). *A guide to criterion-referenced test construction*. Baltimore, John Hopkins.
- Berk, Ronald A. Ed. (1980). *Criterion-referenced measurement: the state of the art*. Baltimore, John Hopkins.
- Cisek, Gregory J. (2001). *Setting Performance Standards. Concepts, Methods and Perspectives*. LEA.
- Educational Measurement: Issues and Practice*. (1994). Número especial sobre pruebas referidas a criterio. Vol. 13, N° 4.

- Popham, W. James (1978). *Criterion-referenced measurement*. Englewood Cliffs, Prentice-Hall.
- Shrock, S. A. y W. C. Coscarelli (1989). *Criterion-referenced test measurement*. Reading, Addison Wesley.
- Pruebas adaptativas, computarizadas, de ejecución y otros desarrollos
- Anderson, Lorin W. (2000). *Assessing Affective Characteristics in the Schools*. LEA.
- Bennett, R. E. y W. C. Ward, Eds. (1993). *Construction versus choice in cognitive measurement: Issues in constructed response, performance testing and portfolio assessment*. Lawrence Erlbaum.
- Brannick, Michael, E. Salas y C. Prince (1997). *Team Performance assessment and Measurement*. LEA.
- Cizek, Gregory J. (1999). *Cheating on Tests: How to Do It, Detect It, and Prevent It*. LEA.
- Drasgow, Fritz y J. B. Olson-Buchanan (1999). *Innovations in Computerized Assessment*. LEA.
- Frary, Robert B. Y George H. Olson (1985). *Detection of Coaching and Answer Copying on Standardized Tests*. Virginia Polytechnic Inst. and State Univ. 9 pp. ED 262 057.
- Glass, Cees A. W. (1998). *Quality Control of On-Line Calibration in Computerized Assessment*. Research Report 98-03. Enschede, Twente University, 24 pp. ED 421 549.
- Hambleton, Ronald, K. (1993). *Translating Achievement Tests for Use in Cross-National Studies*. 32 pp. 358 128.
- Kane, Mitchell Y R. Mitchell Eds. (1996). *Implementing Performance Assessment*. LEA.
- Khatttri, Nidhi, Alison L. Reeve Y Michael B. Kane (1998). *Principles and Practices of Performance Assessment*. Lawrence Erlbaum Assoc.
- Madaus, G. F. (1994). A technological and historical consideration of equity issues associated with proposals to change the nation's testing policy. *Harvard Ed. Review*. 64 (1) pp. 76-95.
- Mills, Craig N. et al. Eds. (2002). *Computer Based Testing*. Lawrence Erlbaum.
- Nichols P.D. (1994) A Framework for Developing Cognitively Diagnostic Assessment. *Review of Educational Research*. Vol. 64. N° 4:575-603.
- Sands, William A., B. K. Walters y J. R. McBRIDE (1997). *Computerized Adaptive Testing: From Inquiry to Operation*. ASC.
- Wainer, Howard et al. Eds. (2000). *Computerized Adaptive Testing. A Primer*. LEA.
- Weiss, David J. (1983). *New Horizons in Testing: Latent Trait Test Theory and Computerized Adaptive Testing*. ASC.
- Wilcox, Sandra K. Y P. E. Lanier (2000). *Using Assessment to Reshape Mathematics Teaching*. LEA.
- Wolf D., J. Bixby Y H. Gardner. (1991) To use their Minds Well: Investigating New Forms of Student assessment. *Review of Research in Education*. Vol.17 pp. 31-74.
- Estándares de calidad de pruebas y procedimientos de metaevaluación
- AERA-APA-NCME (1999). *Standards for educational & psychological testing*. Washington AERA.
- AAHE Assessment Forum (1992). *Principles of Good Practice for Assessing Student Learning*. Washington, American Association for Higher Education.
- Brown, Frederick G. (1980). *Guidelines for test use: A commentary on the Standards for Educational and Psychological Test*. Nat. Council on Measurement in Education.
- Consejo Asesor Externo (2000). *Estándares de calidad para instrumentos de evaluación en gran escala*. México, Ceneval.
- Educational Testing Service (2000). *ETS Standards for Quality and Fairness*. Princeton.
- Martin, Michael et al. (1999). *Technical Standards for IEA Studies*. Amsterdam, IEA.
- Linn, Robert L. et al. (1989). *Quality of Standard Tests. Final Report. School Reform Assessment Project*. Los Angeles, UCLA-CRESST, 155 pp. ED 334 205.
- Martínez Rizo, Felipe et al. (2001). *Estándares de calidad para instrumentos de evaluación educativa*. México, Ceneval.
- McLean, Leslie D. (1985). *The Craft of Student Evaluation in Canada: A Report to the CEA on Policies, Practices and Uses of Assessments of Student Achievement*. Toronto, Canadian Education Association, 63 pp. ED 268 159.

- Nitko, Anthony J. Y Allan Pettie (1989). *The Sixteen Quality Indicators: Standards for Evaluating Criterion-Referenced Tests*. Annual Meeting of the AERA, 19 pp. ED 306 293.
- Ory, John C. (1992). *Meta-Assessment: Evaluating Assessment Activities*. *Research in Higher Education*. Vol. 33 N° 4.
- Shimberg, Benjamin (1995). *Assuring the Quality of Licensing and Certification Programs*. 11 pp. Annual Meeting of the NCME. ED392848
- The Joint Committee On Standards For Educational Evaluation. Propriety Standards. 1981. In *Assessment & Program Evaluation*, pp. 465-483.
- The Joint Committee On Standards For Ed. Evaluation. (1981). *Standards for Evaluation of Educational Programas, Projects and Materials*. Traducida al Español, México, Trillas.
- The Joint Committee On Testing Practices. (1989). Code of Fair Testing Practices in Education. APA. In *Assessment & Program Evaluation*, pp. 485-488.
- Trochim, William M. K. Y Ronald J. Visco (1986). Assuring Quality in Education Evaluation. *Educational Evaluation and Policy Analysis*. Vol. 8, No. 3, pp. 267-276.
- Western Association Of Schools & Colleges (1992). *Principles Of Good Practice in Assessment. Achieving Institutional Effectiveness Trough Assessment*. Oakland, WASC.

6. Evaluaciones en gran escala

6.1. Generalidades y metodología

- David R.G. (1990) Educational System Assessment and Planning Models. En Walberg-Haertel, 1990:710-714.
- Dhingra R. (1991) *Improving the information system for planning the quality of primary education: the case of India*. Monograph. N° 5. París, IIPE.
- Grisay A. Y L. Mahlack. (1991) *The quality of education in developing countries: a review of some research studies and policy documents*. París, IIPE.
- Martínez Rizo, Felipe (1996). *La calidad de la educación en Aguascalientes. Diseño de un sistema de monitoreo*. Aguascalientes. UAA-IEA.
- Keeves John P. (1994) *National examinations: design procedures and reporting* París UNESCO-IIPE. Colección Fundamentals of Educational Planning No. 40.
- Klein, Stephen P. Y Laura Hamilton (1999). *Large-Scale Testing. Current Practices and New Directions*. Santa Monica, Rand Education.
- Livingstone I. D. (1990) Monitoring National Standards. En Walberg H. J. y Haertel G. D. (Eds.) (1990), pp. 435-440.
- OECD (1995) *OECD Education Statistics 1985-1992*. París. OECD.
- Phelps, Richard P. (2000). Trends in Large-Scale Testing Outside the United States. *Educational Measurement: Issues and Practice* (Spring), pp. 11-21.
- Puryear J.M. (1993) *Status and Problems of International Educational Statistics and Research*. CIES Newsletter, No. 103: 4.
- Ross K.N. (1992) Diseño de muestra para estudios internacionales de rendimiento escolar. *Perspectivas*. Vol. XXII, N° 3, pp. 327-339.
- Ross, Kenneth N. Y L. Mählick (1990). *Planning the quality of education. The collection and use of data for informed decision-making*. Oxford, Pergamon-IIPE.
- Ross K.N. Y T.N. Postlethwaite. (1990). *Planning the Quality of Education*, in Walberg-Haertel, 1990: 702-710.
- Taylor C.(1994). Assessment for Measurement or Standards: The Peril an Promise of Large-Scale Assessment Reform. *American Educational Research Journal*. Vol. 31. N° 2:231-262.
- Thelot C.(1993). *L'évaluation du système éducatif*. París. Nathan.
- Tindal, Gerald Y Thomas M. Haladyna (2002). *Large-Scale Assessment Programs for all Students*. Mahwah, NJ, Lawrence Erlbaum Assoc.
- *Wall Chart*, Washington, US Department of Education, annual.

6.2. Experiencias particulares

6.2.1. Sobre el College Board

- Angoff, William H. Ed. (1971). *The College Board Admissions Testing Program*. Nueva York, College Entrance Examination Board.
- Brigham, Carl C. et al. (1926). The Scholastic Aptitude Test of the College Entrance Examination Board. En Fiske, Thomas S. Ed. (1926). *The Work of the College Entrance Examination Board 1901-1925*. Nueva York, Ginn & Co. Págs. 44-63.
- Donlon, Thomas F. Ed. (1984). *The College Board Technical Handbook for the Scholastic Aptitude Test*



and Achievement Tests. Nueva York, College Entrance Examination Board. Especialmente los capítulos I (The Admissions Testing Program: A Historical Overview, pp. 1-11) y II (Psychometric Methods Used in the Admissions Testing Program, pp. 13-36).

Kellaghan Th. Ed. (1995). *Admission to Higher Education. Issues and Practice*. Educational Research Center. IAEA. Princeton, N.J.

6.2.2. Sobre el NAEP

Ballator, Nada (1996). *The NAEP Guide: A Description of the Content and Methods of the 1994 & 1996 Assessments*. Princeton, N. J. Educational Testing Service. 95 pp. ED 401341. For the 1990 and 1992 assessments ED 325 504.

O'Sullivan, Christine (1995). *The Cost of Performance Assessment in Science: The NAEP Perspective*. Princeton, ETS. 14 pp. Paper presented at the Annual Meeting of the NCME ED 384 638.

Walberg H. J. (1990). National Assessment of Educational Progress: Retrospect and Prospect. En Walberg-Haertel 1990:435-440.

6.2.3. Sobre la IEA.

Husén, Torsten y T. Neville Postlethwaite (1996). A Brief History of the International Association for the Evaluation of Educational Achievement (IEA). *Assessment in Education: principles, policy & practice*. Vol. 3 (Julio) No. 2, pp. 129-141.

Martin, Michael O., E Ina V. S. Mullis Eds. (1996). *Third International Mathematics and Science Study: Quality Assurance in Data Collection*. Center for the Study of Testing, Evaluation, and Educational Policy, Boston College, e International Association for the Evaluation of Educational Achievement. 286 pp. ED 406 417

---- y Dana L. Kelly Eds. (1996). *Third International Mathematics and Science Study. Technical Report. Volume I: Design and Development*. IEA. 270 pp. ED 406 418.

Postlethwaite T.N. International Association for the Evaluation of Educational Achievement (IEA). En Husen-Postlethwaite, 1985: 2645-2646.

6.2.4. Otras experiencias

Álvarez Cárdenas, Rosa (2000). La evaluación del sistema educativo cubano. Una experiencia. *Revista de Educación*, No. 321, pp. 41-46.

Álvaro Page M. et al. (1988) *Evaluación externa de la reforma experimental de las enseñanzas medias*. Madrid, Centro de Investigación y Documentación Educativa.

Ángulo Rasco J. F. (1994). Anotaciones al INCE. *Cuadernos de Pedagogía*. N° 219:33-39.

Baker E.L. Y B.H. Choppin. (1990). Minimum Competency Testing. En Walberg-Haertel 1990:499-502.

Barrera, Susana (2000). Estrategias para la viabilidad de un sistema de evaluación de la calidad de la educación: caso Bolivia. *Revista de Educación*, No. 321, pp. 97-114.

Brackenbury, Jim (2000). Canadá: informe de evaluación nacional. *Revista de Educación*, No. 321, pp. 47-60.

Carron, Gabriel y T. G. Chau (1992) *Implicaciones del estudio internacional sobre el mejoramiento de servicios de educación primaria para la planificación y el monitoreo*. Paris. IIEP.

Cosgrove, Judith (2000). El sistema nacional de evaluación de resultados educativos: Irlanda. *Revista de Educación*, No. 321, pp. 23-34.

Guimaraes de Castro, María Helena et al. (2000). El sistema de evaluación de la educación básica de Brasil. *Revista de Educación*, No. 321, pp. 61-80.

Martínez Rodríguez, Félix Francisco (2000). El Sistema Nacional de Evaluación Educativa de México. *Revista de Educación*, No. 321, pp. 35-40.

Ministerio de Educación (1991) *Programa de mejoramiento de la calidad en escuelas básicas de sectores pobres*. Santiago de Chile. Ministerio de Educación.

Ministerio de Educación. (1991) *SIMCE 1991. Manual de interpretación de resultados*. Santiago de Chile, M. de Educación.

- OCDE-PISA (2000). *Measuring Student Knowledge and Skills. The PISA 2000 Assessment of Reading, Mathematical and Scientific Literacy*. París, OCDE.
- De la Orden Hoz, Arturo *et al.* (1999). *Los resultados escolares. Diagnóstico del Sistema Educativo, 1997. La escuela secundaria obligatoria*. Madrid, MEC-INCE.
- OREALC (1994) *Medición de la calidad de la educación*. UNESCO-OREALC. Santiago de Chile.
- Vol. I: ¿Por qué, cómo y para qué? Vol. II: Instrumentos. Vol. III: Resultados de siete países.
- Puente Azcutia, Julio (2000). La evaluación del sistema educativo español. *Revista de Educación*, No. 321, pp. 81-96.
- Ross K.N. (1994) *The Establishment of a Southern Africa Consortium for the Monitoring of the Quality of Education*. París, IIEP.
- Sacmeq. (1995). *Southern Africa Consortium for Monitoring Educational Quality*. París, IIEP.
- Secretaría de Educación Pública (2000). *Distribución de los planteles públicos de educación primaria y secundaria según el nivel de aciertos de sus alumnos en los exámenes de carrera magisterial*. México, SEP.
- Tiana Ferrer, Alejandro (2000). *Cooperación internacional en evaluación de la educación en América Latina y el Caribe. Análisis de la situación y propuestas de actuación*. Washington, Banco Interamericano de Desarrollo.
- Varios (1991). Antología de textos sobre competencias mínimas. En Latapí, Pablo (Coord.) 1991. *Educación y escuela. Lecturas básicas para investigadores de la ed. II Aprendizaje y rendimiento*. México, Nueva Imagen, pp. 135-202.
- Vari, Peter (2000). La evaluación educativa Monitor y el sistema de evaluación de Hungría. *Revista de Educación*, No. 321, pp. 9-22.
- Wolff, Lawrence (1998). *Las evaluaciones educacionales en América Latina: Avance actual y futuros desafíos*. Santiago de Chile, PREAL.

7. Indicadores

7.1. Indicadores sociales

- Delors, Jacques Ed. (1971) *Contributions a une recherche sur les indicateurs sociaux*. París: Futuribles, SEDEIS.
- Departamento de Asuntos Económicos y Sociales (1975). *Hacia un sistema de estadísticas sociales y demográficas*. Nueva York: ONU.
- Russett, Bruce M., H. R. Alker, K. W. Deutsch Y H. D. Lasswell (1965). *World Handbook of Political and Social Indicators*. Nueva Haven: Yale Univ. Press. Traducción al español Eura-mérica, 1968.
- Zemelman, Hugo (1989). *Crítica epistemológica de los indicadores*. México: Colmex.

7.2. Indicadores educativos en general

- Banta, Trudy W. (1991) *Education Counts. Assessment Update*. Vol.3. N° 5:3-4.
- Bottani N. Y Delfau I. (1990) The Search for International Educational Indicators. *The OECD Observer*. N° 162: 14-18.
- Bryk Anthony y K.L. Hermanson. (1992) Educational Indicator System: Observations on their Structure, Interpretation and Use. *Review of Research in Education*. Vol. 19: 451-484.
- Burstein et al. (1992) Education Indicators. En Alkin, 1992, 2: 410-418.
- CERI-INES (1995^a) *Education at a Glance: OECD Indicators*. París, OECD.
- CERI-INES (1995b) *Measuring the Quality of Schools*. París: OCDE. ED 411 321, 238 pp.
- CERI-INES (1995c) *Measuring What Students Learn*. París: OCDE.
- CERI-INES (1995d) *Public expectations of the final stage of compulsory education*. París: OECD.
- CERI-INES (1991) *An International Handbook of Educational Indicators*. París: OCDE.

- CERI-INES (1991) *General Assembly on International Education Indicators Handbook*. París: OCDE.
- CERI-INES (1992). *The OECD International Education Indicators: A Framework for Analysis*. París: OCDE.
- CERI-INES (1973). *Indicators of Performance of Educational Systems*. París: OCDE.
- David J.L. (1987) *Improving Education with Locally Developed Indicators*. New Brunswick. Center for Policy Research in Education-Rutgers University/The Rand Corporation/Wisconsin Center for Education Research.
- Emin J.C. (1995) *La mise en place d'un dispositif d'indicateurs pour le pilotage des établissements secondaires français*. En CERI-INES 1995b, Cap. 10 pp. 201-231.
- Hopkins D. (1994) Process Indicators for School Improvement. En Tuijnman-Bottani, Cap. 8:145-170.
- Hopkins, D. (1990) International Indicators and Evaluation of Educational Systems. *INES News*, N° 4: 1-7.
- Johnstone J. N. (1981) *Indicators of Education Systems*, Londres, Kogan Page-UNESCO.
- Kaagan S.S. Y R.J. Loby. (1989) *State Education Indicators: Measured Strides, Missing Steps* New Brunswick. Rutgers Univ. Center for Policy Research in Education.
- KIRST. (1989) en Tuijnman-Bottani 1994:24-25.
- (s/f) *Laboratorio Latinoamericano de Evaluación Educativa*. Mimeo.
- De Landsheere G. (1994). *Le pilotage des systèmes d'éducation*. Bruselas. De Boeck.
- (1994) *L'Etat de l'école*. París: MEN.
- Maguire Th. O. (1993) Some Thoughts on the Alberta Version of Educational Quality Indicators. *The Alberta Journal of Educational Research*. Vol. XXXIX No. 2, pp. 275-277.

- McEwen N. (1993) Educational Quality Indicators. *The Alberta Journal of Educational Research*. Vol. XXXIX. No. 2:167-177.
- Ross K. N. Y T. N. Postlethwaite. (1992) *Indicators of the quality of education: a summary of a national study of primary schools in Zimbabwe*. París: UNESCO-IIPE.
- Rumberger R.W. (1994) Labour Market Outcomes as Indicators of Educational Performance. En Tuijnman-Bottani, 1994:265-286.
- Sauvageot C. (Ed.) (1992) *Primary Education in Lesotho. Indicators 1992*. París, IIPE.
- Sauvageot C. (Ed.) (1993) *L'enseignement fondamental au Mali. Indicateurs 1993*. París: IIPE.
- Special Study Panel on Education Indicators. (1991) *Education Counts. An Indicator System to Monitor the Nation's Educational Health*. Washington. National Center for Educational Statistics. USA Dpt. of Education.
- Tuijnman A. Y N. Bottani (Eds.) (1994) *Making Education Count. Developing and Using International Indicators*. París: OCDE.
- West A.R., P.H. Wait Y S. Thomas. (1994) *Community Educational Indicators Phase 2. Report to M. States*. Consejo de Europa.
- Wyatt T. (1994). Education indicators: a review of the literature. En Tuijnman y Bottani, 1994:99-121.
- 7.3. Indicadores sobre educación superior**
- Association of Universities and Colleges of Canada (1995). A Primer on Performance Indicators. *Research File*. Vol. 1 (Jun) No. 2, 10 pp. Ottawa, ED 386 100
- Barton, Paul E. Y A. Lapointe (1995). *Learning by Degrees: Indicators of Performance in Higher Education*. Princeto, ETS, 84 pp. ED 379 323.
- Birch, Derek W. Y R. E. Cuthbert (1979). *Resource Utilisation Performance Indicators in the Public Sector of Higher Education, or Never Mind the Technique Feel the Structure*. Annual Conference of the SRHE, ED 248 931, 20 pp.
- Cave, Martin, St. Hanney, M. Henkel Y M. Kogan (1997). *The Use of Performance Indicators in Higher Education: The Challenge of the Quality Movement*. 3d Edition. Higher Education Policy Series 34. ED 415 782, 280 pp. 2nd Edition, ED334951, 192 pp. 1st ed. ED 328 198, 126 pp.
- Christal, Melodie E. Y D. P. Jones (1985). *A Common Language for Postsecondary Accreditation: Categories and Definitions for Data Collection*. Boulder-Washington, NCHEMS-The Council on Postsecondary Accreditation.
- David, Dorothy (1992) *The Real World of Performance Indicators. A Review of Their Use in Selected Commonwealth Countries*. Londres: C. H. Ed. Mgmt. Service. ED 416 815, 44 pp.
- Foster, Pablo (1991). *Performance Indicators and Adult Education*. Blagdon: Further Education Staff Coll. ED 341 848, 32 pp.
- Gaither, Gerald, B. P. Nedwek Y J. E. Neal (1994). *Measuring Up: The Promises and Pitfalls of Performance Indicators in Higher Education*. ASHE-ERIC Higher Education Reports No. 5. Wasington, George Washington University, 159 pp. ED 383 279.
- Kells, H. R., Ed. (1990). *The Development of Performance Indicators for Higher Education: A Compendium for Eleven Countries*. París: OCDE-IMHE, ED331355, 134 pp.
- Nedwek, Brian P. Y J. E. Neal (1993). *Performance Indicators and Rational Management Tools: A Comparative Assessment of Projects in North America and Europe*. Documento presentado al Foro Anual de AIR ED360917, 43 pp.
- Ruppert, Sandra S. Ed. (1994). *Charting Higher Education Accountability. A Sourcebook on State-Level Performance Indicators*. Denver: Education Commission of the States.
- Taylor, Barbara E., J. W. Meyerson Y W. F. Massy (1993). *Strategic Indicators for Higher Education. Improving Performance*. Princeton, N. J., Peterson's Guides.
- (1996). *Strategic Indicators for Higher Education: Vital Benchmarks and Information to Help You Evaluate and Improve Your Institution's Performance*. Princeton, Peterson's.
- Yorke, Mantz (1997). *Can Performance Indicators Be Trusted?* Documento presentado al Foro de ED 418 660, 43 pp.

8. Evaluación de escuelas

- Abbott R., S. Steadman Y M. Birchenough. (1988) *GRIDS Primary School Handbook*. Nueva York: Longman.
- Aubegny J. (1995). *The role of evaluation and accountability as a management tool and a means of improving effectiveness at school and local authority levels*. París: OCDE.
- Austin y H. Garber (1985) *Research on exemplary schools*. Orlando. Academic Press.
- Barbera Albalat V. (1990). *Método para evaluación de centros*. Madrid: Escuela Española.
- Beare H., B.J. Caldwell Y R.H. Milligan (1989). *Creating an Excellent School. Some New Management Techniques*. Londres. Routledge.
- Blanco, Felipe L. (1995). Modelo para una autoevaluación modular de una institución educativa. *Revista de Investigación Educativa*. N° 21, pp. 47-66.
- Blum R. Y J. A. Butler Eds. (1989). *School-Leader Development for School Improvement*. París: OCDE-ISIP.
- Bolam R. (1985) School Inspectors and Advisers. En Husen-Postlethwaite, 1985: 4429-4432.
- Bollen R. Y D. Hopkins Eds. (1987). *School Based Review: Towards a Praxis*. International School Improvement Project, OCDE-ISIP. Leuven, ACCO.
- Casanova R. Ma. A. et al. (1991) *La Evaluación del Centro Educativo*. Curso de Formación para Equipos Directivos. Unidad Temática 5. Madrid. MEC.
- Climaco C. Y S. Araujo. (1995) The Effectiveness of Schools and Educational Resource Management. National Report. Portugal: Lisboa. Ministerio de Educação.
- De Caluwé, L., E. C. H. Marx Y M. W. Petri (1988). *School Development: Models and Change*. París: OCDE-ISIP.
- Deelsa. (1992) *The role of the inspectorate and inspectors in the development and monitoring of school management and effectiveness*. The effectiveness of schooling and of educational resource management. Monograph. N° 2. París: OCDE.
- Electronic Industries Foundation (1996). *National Skills Standards Development Project. Study of the State of the Art of Certification and Accreditation Programs*. Washington, 29 pp. ED 396 133. Related documents ED 389 946.
- Fueyo, D. Y R. Fernández (1989). Análisis de la documentación sobre evaluación de centros en España. *Revista de Investigación Educativa*. Vol. 7, N° 13, pp. 183-199.
- Gairin Sallan J. (1993) *La autoevaluación institucional como vía para mejorar los centros educativos*. Bordon. Vol. 45. N° 3.
- Gairin Sallan J. (1995) La evaluación del proyecto educativo de centro. *Organización y gestión educativas*. N° 1:31-36.
- Hallak J. (1992) Managing schools for educational quality and equity: finding the proper mix to make it work. Contributions N° 8. París. IIPE.
- Hameyer U. Y S. Loucks-Horsley Eds. (1989). *New Technologies and School Improvement, Support Policies and Practices*. París: OCDE.
- Hopes, C. Ed. (1986). *The School Leader and School Improvement: Case Studies from Ten OCDE Countries*. París: OCDE-ISIP.
- Hopkins, D. (1994). Process indicators for school improvement. En Tuijn Many Bottani, 1994: cap. 8.
- Ed. (1989). *Doing School Based R: Instruments and Guidelines*. París: OCDE-ISIP.
- Ed. (1985). *School Based Review for School Improvement*. París: OCDE-ISIP.



- Hunt Ogden E. Y V. Germinario. (1994) The Nation's Best Schools. Blueprints for Excellence. Vol. 1: Elementary and Middle Schools Lancaster. Technomic Publ.
- Lecoïnte M. Y M. Rebinguet (1990). *L'audit de l'établissement scolaire*. París: Les éditions d'organization.
- Levin, Henry M. (2000). *Las Escuelas Aceleradas: Una Década de Evolución*. Washington-Santiago: PREAL.
- Lillis K.M. (1992) *Improving basic education: preconditions for successful inspection and supervision. Implications for training. Increasing and improving the quality of basic education*. Monograph N°7. París: IIPE.
- López J.A. Y P.Darder (1985). *Quafe-80. Cuestionario para el análisis del funcionamiento de la escuela*. Barcelona: Onda.
- Madaus G., P. Airasian Y T. Kellaghan (1980). *School effectiveness: a reassessment of the evidence*. Nueva York: McGraw-Hill.
- Mestres, J. (1989). Gestión y evaluación de centros escolares. *Apuntes de Educación (dirección y administración)* N° 33 (abril-junio), pp. 2-4.
- Miles M. B., M. Ekholm Y R. Vandenberghe Eds. (1987). *Lasting School Improvement: Exploring the Process of Institutionalization*. OCDE-ISIP. Leuven: ACCO.
- Radnor, Hilary (1993). *Moderation and Assessment Project-South West. A Presentation of a Model for Moderating Pupils' Work That Is Teacher Assessed*. Annual Meeting of the AERA, 17 pp. ED 361 397.
- Rodríguez Díaz, B. (1993). Modelo de evaluación externa de la eficacia de los centros educativos. *Revista de Investigación Educativa*. N° 21, pp. 7-27.
- Sabiron Sierra F. (1990) *Evaluación de Centros Docentes. Modelo, Aplicaciones y Guía*. Zaragoza: Central de Ediciones.
- Saez M.J. Y M.Fernández (1993). La evaluación del proyecto educativo de un centro: el problema de los indicadores. *Revista de Ciencias de la Educación*. N° 153, pp. 23-36.
- Santos Guerra M.A. (1993) *Estrategias para la evaluación interna de los Centros Educativos*. Madrid: MEC.
- Seashore Louis K. Y S. Loucks-Horsley Eds. (1990). *Supporting School Improvement: A Comparative Perspective*. OCDE-ISIP. Leuven: ACCO.
- Stëgo, N. E. et al. Eds. (1987). *The Role of School Leaders in School Improvement*. OCDE-ISIP. Leuven: ACCO.
- Tiana A. Y Municio P. (1995) La evaluación de centros en el INCE. Entrevista con Alejandro Tiana Farrer por Pedro Municio. *Organización y gestión educativa*. N° 2.
- Van Den Berg R. M. Y R. Vandenberghe Eds. (1986). *Strategies for Large-scale Change in Education: Dilemmas and Solutions*. OCDE-ISIP: Leuven, ACCO.
- Van Velzen, W. G. et al. (1985). *Making School Improvement Work: A Conceptual Guide to Practice*. International School Improvement Project, OCDE-ISIP. Leuven: ACCO.
- Varios (1991). *La evaluación del centro educativo*. Madrid: MEC.
- Wick, John W. (1987). *School-Based Evaluation*. Dordrecht: Kluwer.