EVALUACIÓN Y REFORMA EDUCATIVA
Opciones de política

Benjamín Álvarez H.
Mónica Ruiz-Casares

Editores

1997

El Proyecto para el Avance de la Educación Básica y la Alfabetización (ABEL) es gestionado por la Academia para el Desarrollo Educativo (AED) juntamente con Creative Associates International, Inc., Education Development Center, Florida State University, Harvard Institute for International Development, Juarez and Associates, y el Research Triangle Institute. El Proyecto ABEL está financiado por la Agencia para el Desarrollo Internacional de los Estados Unidos (USAID).

Este documento ha sido elaborado por la Academia para el Desarrollo Educativo con patrocinio y orientación de la Oficina para América Latina y el Caribe y del Centro para el Desarrollo de la Capacidad Humana de la USAID. Los hallazgos, conclusiones y recomendaciones presentados en los documentos de ABEL son los de sus autores y no reflejan las opiniones de ninguna de las instituciones asociadas con el Proyecto ABEL ni de la Agencia para el Desarrollo Internacional de los Estados Unidos.

Permitida la reproducción de los datos contenidos en esta publicación, siempre que se cite su procedencia.

Proyecto 936-5832
Contrato HNE-Q-00-94-00076-00

CONTENIDOS

PREFACIO vii

INTRODUCCIÓN ix

1. CONTRIBUCIÓN DE LA EVALUACIÓN AL APRENDIZAJE DE LAS NACIONES
Benjamín Álvarez H. 1
EVALUACIÓN DEL APRENDIZAJE DE LAS NACIONES, LAS ORGANIZACIONES Y LOS INDIVIDUOS..... 1
DIMENSIONES Y FUNCIONES DE LA EVALUACIÓN 4
EVOLUCIÓN DE LA PRÁCTICA DE LA EVALUACIÓN EDUCATIVA EN EL HEMISFERIO 8
ÁREAS CRÍTICAS DE LA EVALUACIÓN DE LA EDUCACIÓN 11
CRITERIOS PARA UNA AGENDA DE POLÍTICAS DE EVALUACIÓN 18

Parte I
SISTEMAS NACIONALES

2. SEGUIMIENTO DE LOS RESULTADOS EDUCATIVOS NACIONALES
Thomas Kellaghan 23
INTRODUCCIÓN..... 23
FUNDAMENTO DE LAS EVALUACIONES NACIONALES 25
INDICADORES 26
USOS DE LOS DATOS DE UNA EVALUACIÓN NACIONAL 28
DESCRIPCIÓN DEL ESCENARIO DE UNA EVALUACIÓN NACIONAL 34
ELECCIÓN DE UNA INSTITUCIÓN PARA REALIZAR LA EVALUACIÓN NACIONAL 35
¿A QUIÉN EVALUAR?..... 37
¿QUÉ SE EVALÚA?..... 42
INSTRUMENTOS DE EVALUACIÓN 44
ADMINISTRACIÓN DE LA EVALUACIÓN NACIONAL 48
TIEMPO Y FRECUENCIA DE LA EVALUACIÓN 50
ANÁLISIS DE LOS DATOS 51
PRESENTACIÓN DE LOS RESULTADOS 52
CÁLCULO DE LOS COSTOS 56
CONCLUSIÓN..... 59

3. EVALUACIÓN Y ESTÁNDARES DE CURRÍCULO
Gilbert A. Valverde 67
INTRODUCCIÓN..... 67
POLÍTICAS EDUCATIVAS, SISTEMAS DE INDICADORES Y EL PROCESO INVISIBLE 69
PROCESO VISIBLE: UN MODELO SISTÉMICO DE OPORTUNIDADES EDUCATIVAS 74
UNA ESTRATEGIA DE EVALUACIÓN SISTÉMICA 81

ALGUNOS EJEMPLOS ADICIONALES DE LA APLICACIÓN DE LAS TÉCNICAS DEL SMSO 88
ALGUNAS CONCLUSIONES: EVALUACIÓN Y FORMULACIÓN DE UNA POLÍTICA DEL CURRÍCULO EFICAZ 95

4. SEGUIMIENTO INTERNACIONAL DE LAS METAS DE DESARROLLO HUMANO: EL CASO DE LA SECRETARÍA PRO TÉMPORE DE LAS AMÉRICAS
María Inés Cuadros 103
ANTECEDENTES 103
LA SECRETARÍA PRO TÉMPORE AMERICANA: 1994- 1996 108
AVANCES EN EL LOGRO DE LAS METAS DE DESARROLLO HUMANO 111
EL ACUERDO DE SANTIAGO 113
CONTRIBUCIÓN DE LA SECRETARÍA PRO TÉMPORE 115

Parte II
LECCIONES DE LA HISTORIA

5. IMPACTO SOCIAL DE LOS SISTEMAS DE EVALUACIÓN DEL RENDIMIENTO ESCOLAR: EL CASO DE CHILE
Erika Himmel 125
INTRODUCCIÓN 125
SISTEMAS NACIONALES DE EVALUACIÓN Y USO DE LA INFORMACIÓN 126

6. SISTEMA DE EVALUACIÓN DE LA CALIDAD DE LA EDUCACIÓN EN COLOMBIA
Gabriel Restrepo 159
INTRODUCCIÓN 159
SISTEMA DE SELECCIÓN: LOS EXÁMENES DE ESTADO PARA EL INGRESO A LA EDUCACIÓN SUPERIOR (1964-1990) 161
SISTEMA NACIONAL DE EVALUACIÓN DE LA CALIDAD: 1991-1997 171
CONCLUSIONES 179

Parte III
EVALUACIÓN Y PROFESIONALISMO DE LOS MAESTROS

7. EVALUACIÓN DE LOS MAESTROS
Carol Dwyer 187
INTRODUCCIÓN 187
EVALUACIONES DE LOS MAESTROS Y MEJORAMIENTO EDUCATIVO 188
METAS, ESTÁNDARES Y EVALUACIONES 188
FINES DE LA EVALUACIÓN 197
DEFINICIÓN DE LA ENSEÑANZA 203
CUESTIONES RELACIONADAS CON LA FORMULACIÓN DE EVALUACIONES DE LOS MAESTROS 207
CONCLUSIÓN 216

8. EVALUACIÓN DE LA ACCIÓN DOCENTE EN LATINOAMÉRICA *Francisco Álvarez Martín, con la colaboración de María José Álvarez y Paula Vergara* 223
INTRODUCCIÓN.....223
¿QUÉ SABEMOS SOBRE LA EVALUACIÓN DE LOS DOCENTES? 225
PRÁCTICA DE LA EVALUACIÓN DOCENTE 239
RECOMENDACIONES..... 253

Parte IV
EVALUACIÓN DE LA ORGANIZACIÓN EDUCATIVA

9. EVALUACIÓN DE LAS ESCUELAS *William J. Webser y Robert L. Mendro* 259
INTRODUCCIÓN.....259
CONCEPTUALIZACIONES ALTERNATIVAS DE EVALUACIÓN 260
CONSIDERACIONES PARA EL DESARROLLO DE UN MODELO DE EVALUACIÓN DE LA ESCUELA..... 264
TRES NIVELES DEL MODELO DE EVALUACIÓN DE LA ESCUELA 276
MODELO DE VALOR AGREGADO DE DALLAS 295
USOS DEL SISTEMA DE VALOR AGREGADO EN LA EVALUACIÓN DE LAS ESCUELAS..... 297
UTILIZACIÓN DE LA INFORMACIÓN DE VALOR AGREGADO 302

10. EVALUACIÓN EN LAS REFORMAS EDUCATIVAS DEL ESTADO DE PARANÁ, BRASIL *María Teresa de la Fuente, Heloisa Luck y Corinna Ramos* 321
INTRODUCCIÓN..... 321
CONTEXTO DE LA REFORMA EDUCATIVA EN BRASIL 321
INICIATIVAS ESTATALES QUE RESPONDEN A LAS DIRECTRICES DE LA REFORMA EDUCATIVA..... 324
EVALUACIÓN NACIONAL DEL RENDIMIENTO ESCOLAR (SAEB) 325
ESTADO DE PARANÁ: EVALUACIÓN DE LAS ACTIVIDADES DE LA REFORMA EDUCATIVA..... 326
SEGUIMIENTO Y EVALUACIÓN DEL RENDIMIENTO ESCOLAR EN PARANÁ 327
SEGUIMIENTO Y EVALUACIÓN DE LA CAPACITACIÓN PERMANENTE DE LOS PROFESIONALES DE LA EDUCACIÓN 330
SEGUIMIENTO Y EVALUACIÓN DE LA GESTIÓN EDUCATIVA: EN BUSCA DE LA EXCELENCIA EN LA ESCUELA 332
UTILIDAD, FACTIBILIDAD Y RELEVANCIA DE LOS PROCEDIMIENTOS DE SEGUIMIENTO Y EVALUACIÓN EN PARANÁ 335
CONCLUSIONES PARA OTROS CONTEXTOS EN BRASIL Y PAÍSES CON DESAFÍOS SIMILARES..... 338

11. CONSTRUCCIÓN DE UN SISTEMA NACIONAL DE EVALUACIÓN EN AGUASCALIENTES, MÉXICO *Margarita María Zorilla Fierro* 345
INTRODUCCIÓN..... 345
EL ESTADO DE AGUASCALIENTES 346
ORIGEN DEL SISTEMA ESTATAL DE EVALUACIÓN DE LA EDUCACIÓN 348
EVALUACIÓN DEL APRENDIZAJE EN 1996 351

ALGUNOS RESULTADOS DE LA MEDICIÓN DEL INDICADOR DE APROVECHAMIENTO ESCOLAR EN LA ESCUELA PRIMARIA 352
EFECTOS DE LAS POLÍTICAS SOBRE EVALUACIÓN 355
LOS RIESGOS Y LAS ALTERNATIVAS 357
ENSEÑANZAS DE LA EXPERIENCIA INICIAL 358

Conclusiones

DILEMAS Y OPCIONES DE POLÍTICA PARA LA EVALUACIÓN EDUCATIVA
Benjamín Álvarez H. y Ray Chesterfield 363
TENDENCIAS PRINCIPALES 363
EJECUCIÓN DE LOS SISTEMAS DE EVALUACIÓN 368
OPCIONES DE POLÍTICA 369
CONCLUSIONES..... 373

AUTORES 375

APÉNDICE: BIBLIOGRAFÍA RECIENTE 377

Prefacio

Los estándares y la evaluación de la educación ocupan un lugar prominente en el diálogo internacional sobre el desarrollo. El catalizador de este diálogo es la necesidad inaplazable que los gobiernos tienen de preparar a todos sus ciudadanos para la sociedad de la información del siglo XXI, liderada por una economía de competencia global. A medida que se examinan más cuidadosamente las implicaciones sociales de la competitividad global, se hace más evidente que mejorar la capacidad de los recursos humanos de las naciones y, por tanto, sus sistemas de educación, constituye una tarea imprescindible. La búsqueda de caminos para mejorar el desempeño de los sistemas educativos conduce inexorablemente a la discusión sobre los estándares y la evaluación.

Aunque la mayor parte de la gente está de acuerdo con la necesidad de mejorar los resultados de la educación en términos del aprendizaje de los estudiantes en los países del hemisferio, no es fácil obtener un consenso sobre la mejor forma de utilización de esta información, o inclusive sobre si los métodos que estamos utilizando miden adecuadamente la eficacia y el desempeño de la educación. El actual debate en torno de la adopción de estándares nacionales de educación en los Estados Unidos constituye un buen ejemplo de las dificultades de la problemática de la evaluación.

Recientemente, se han desarrollado varios proyectos encaminados a mejorar las estadísticas educativas en los países de América Latina. Estos proyectos han aumentado substancialmente la tecnología disponible, pero no han logrado proporcionar orientaciones claras para el desarrollo de estadísticas, indicadores y sistemas de información. La Oficina para América Latina y el Caribe de la Agencia para el Desarrollo Internacional de los Estados Unidos espera que este volumen contribuya a ampliar la base de información y conocimiento que los responsables de las decisiones políticas deben tener para desarrollar sistemas de evaluación y seguimiento, proponer estándares para los servicios educativos y utilizar la información necesaria para mejorarlos. Esperamos que este informe, además de llenar vacíos de conocimiento, promueva nuevas reflexiones, análisis e investigaciones.

En nombre de la Agencia para el Desarrollo Internacional de los Estados Unidos, dedicamos este trabajo a los niños de las escuelas de todo el hemisferio y agradecemos a los autores su valiosa contribución al proyecto. Especialmente agradecemos a la Academia para el Desarrollo Educativo su excelente trabajo y la

dedicación de personas como Francy Hays, Mónica Ruiz-Casares y el Dr. Benjamín Álvarez por su liderazgo intelectual en el examen de temas complejos de las políticas de refoma educativa, estándares y evaluación.

Sarah Wright
Oficina para América Latina y el Caribe
División de Educación y Recursos Humanos
Agencia para el Desarrollo Internacional de los Estados Unidos

Introducción

La educación posee un valor irremplazable en la configuración de la historia de los individuos y las sociedades. Así lo atestiguan los pensadores de todas las épocas quienes han acudido a las analogías predominantes de cada tiempo y lugar para expresar lo que se espera lograr con la actividad de la educación y para reorientar las prácticas predominantes. Aristóteles, por ejemplo, utilizó el prisma de la política para identificar las funciones de la educación en la sociedad. Más recientemente, Piaget se inspiró en la biología para representar los procesos de interacción que ocurren entre el organismo humano y su ambiente y que conducen al aprendizaje. La metáfora preferida hoy para "reinventar" la educación es la del funcionamiento de los mercados, su dinámica de competencia y su eficiencia.

Sin embargo, aunque tales aproximaciones enriquecen el acervo de conocimiento sobre el desarrollo humano y el progreso social y, proporcionan modelos para organizar los sistemas de educación, están inevitablemente constreñidas por sus propios supuestos básicos y no agotan ni las alternativas ni el significado de la educación en la sociedad. "La educación—decía Kant—es un arte cuya práctica debe ser perfeccionada a lo largo de las generaciones....y es el problema mayor y más difícil que pueda plantearse al hombre." Esta limitación natural de nuestros modelos de pensamiento y la incansable preocupación de las sociedades por la educación, reflejan el hecho de que aprender y enseñar son el rasgo característico de la especie humana. Más aún, las sociedades se desarrollan a través de mecanismos de aprendizaje. Es natural entonces, que aprender sobre la educación constituya una tarea permanente de la sociedad. Para juzgar nuestros logros y necesidades nuevas como comunidad de estudiantes y maestros se han establecido los mecanismos de evaluación, a cuyo desarrollo ha contribuido especialmente la ciencia social contemporánea.

Este libro busca contribuir al formidable esfuerzo de renovación educativa en que se encuentran empeñados los países de América Latina y el Caribe, por medio del análisis de las diversas dimensiones de la evaluación y las opciones que de allí surjen para la formulación de la política pública y privada. Su propósito primordial, más que diseñar un modelo o presentar una receta terminada, es el fortalecer las competencias nacionales para recrear continuamente los sistemas educativos, una de las actividades fundamentales del arte de gobernar en la sociedad del conocimiento.

Muchas personas han contribuido a esta publicación, entre otras Jim Hoxeng de la Agencia para el Desarrollo Internacional de los Estados Unidos (USAID), quien

leyó cuidadosa y críticamente todos los trabajos y Sarah Wright, quien participó en todas las actividades de planificación y ejecución de los estudios que sirvieron de base. Francy Hays de la Academia para el Desarrollo Educativo, administró el proyecto y proporcionó invaluable consejo y liderazgo. Los autores y editores agradecen también su contribución a los traductores y revisores de los manuscritos, particularmente a Ray Chesterfield de Juarez y Asociados y a Ward Henevel del Banco Mundial. La difusión inicial de este trabajo, fue posible gracias al Banco Mundial y al Programa PREAL, que junto con la USAID patrocinaron la realización de un taller internacional sobre el tema

CAPÍTULO 1

CONTRIBUCIÓN DE LA EVALUACIÓN AL APRENDIZAJE DE LAS NACIONES

Benjamín Álvarez H.

En pocos temas de política social se ha dado un acuerdo tan amplio a través del hemisferio americano como en torno de la necesidad de renovar —y en la mayor parte de los casos reformar— los sistemas nacionales de educación. Con la esperanza de resolver problemas antiguos y de lograr una inserción definitiva en la sociedad global, los países de América Latina y el Caribe están promulgando nuevas políticas educativas que se caracterizan por:

- destacar los resultados del proceso educativo más que sus insumos
- abrir oportunidades para que la comunidad y la sociedad civil se comprometan con la educación
- proporcionar mayor autonomía a las escuelas
- promover la calidad del sistema y su eficiencia
- mejorar la competencia profesional de los educadores y
- fomentar una mayor igualdad de oportunidades

A medida que se van ejecutando las reformas educativas, el papel de los gobiernos se transforma de administrador a evaluador y decisor de política. Y a medida que más personas y organizaciones participan activamente en la educación, las prácticas escolares y los resultados obtenidos por los estudiantes se convierten en objeto de mayor atención. La importancia del aprendizaje social promovido por el establecimiento de una cultura y unos procedimientos de evaluación seguramente aumentará, motivada por necesidades tanto políticas como prácticas. Este libro introduce al lector a la problemática de la evaluación de la educación en un clima de reforma, sus temas críticos para la decisión política y su imprescindible contribución al aprendizaje de las naciones.

EVALUACIÓN DEL APRENDIZAJE DE LAS NACIONES, LAS ORGANIZACIONES Y LOS INDIVIDUOS

Desde tiempos remotos, las sociedades han considerado la educación unas veces como un medio de reproducción de valores y conocimientos y otras como un

vehículo de movilidad y cambio social, pero nunca antes habían imaginado que su destino y lugar en el concierto mundial de las naciones dependiera de su capacidad de aprendizaje y que "la batalla por elevar el nivel de vida de los pueblos se luchara primero y fundamentalmente en las aulas escolares"(The Economist, 1997).

Las escuelas más antiguas sobre las que poseemos alguna información, tales como las de la civilización sumeria, tenían la misión de prestar atención prioritaria a las necesidades de grupos selectos de la sociedad, por ejemplo, los escribas y funcionarios gubernamentales. Y así continuaron por siglos. Hoy día, sin embargo, la educación dejó de ser un asunto de responsabilidad, motivación y vocación individual para convertirse en un tema de política pública y en un requisito indispensable para la supervivencia de las naciones.

El comercio, la salud, el trabajo y el bienestar humano dependen de la inserción que tengan las personas y los países en la densa e intrincada red de información y conocimiento que envuelve el mundo entero y que se encuentra en permanente renovación. Por primera vez en la historia la interacción humana puede realizarse de forma simultánea y global, de tal manera que todos los pueblos, en mayor o menor grado, pero sin excepción alguna, se encuentran envueltos en una aventura de aprendizaje sin fronteras. Los más optimistas visionarios de la evolución social del pasado reciente se sorprenderían ante las oportunidades que el espacio cibernético y la cada día más estrecha relación entre el conocimiento y la economía abren a la simultánea diversidad y convergencia universal del pensamiento.[1] A medida que el mundo se integra más, como resultado de esa convergencia, la característica que diferenciará unas regiones de otras será la calidad de sus instituciones públicas. Las áreas de mayor éxito serán quizás las que posean los sistemas más competentes y eficientes para apoyar los intereses colectivos, especialmente la producción de nuevas ideas (Romer, 1993). El desarrollo de tales sistemas depende a su vez de la capacidad y las oportunidades de aprendizaje existentes en una sociedad.

Aprender a aprender no constituye entonces una meta exclusivamente individual, sino también una necesidad para las naciones. Las naciones, como los individuos, evolucionan y cambian; interactúan con su ambiente a través de procesos de asimilación y adaptación. En cierto sentido se puede decir que aprenden. Pero sin mecanismos para repensar y revisar sus prácticas educativas, para conectar sus ideales y utopías con el conocimiento científico y comparar sus logros, los sistemas educativos nacionales pierden su norte y rompen los vínculos con su propia identidad. El problema ya no consiste solamente en desarrollar contextos especialmente diseñados para el aprendizaje de todos en una sociedad —tales como las escuelas, los centros de capacitación, los programas de educación no-formal— sino además en dotar tales contextos de una capacidad de asimilación y adaptación constantes y de búsqueda de óptimos resultados; resultados que aunque intangibles, son la base del bienestar y progreso de los países, de la productividad de las empresas y del futuro de los individuos (Porter, 1990; Reich, 1990). Esta es una de las razones que explican el creciente interés político en torno del tema de la

evaluación.

Este interés de las naciones, las organizaciones y los individuos por mejorar sus oportunidades de aprendizaje, se manifiesta en varias tendencias convergentes. La primera, y, quizás más pública expresión de estas tendencias, la constituye la multiplicación reciente de las reformas de los sistemas nacionales de educación, basadas en la participación ciudadana y orientadas al logro de resultados de alta calidad. La segunda, es la búsqueda incansable de nuevos modelos de organización del mundo productivo para adaptarse continuamente a la dinámica de la economía del conocimiento. La tercera, es el incremento de las necesidades de acreditación del aprendizaje que las personas adquieren hoy día en múltiples contextos a lo largo de toda su vida.

Las reformas de la educación — con las cuales la mayoría de los países del mundo esperan inaugurar el próximo milenio y responder a las demandas de la nueva economía— representan una oportunidad de progreso a condición de que los esfuerzos de renovación estén acompañados de un sistema inteligente de reflexión, seguimiento, reajuste y reinvención. La historia de las reformas educativas sugiere que no existe un algorítmo o fórmula mágica que pueda reemplazar la capacidad nacional y local para aprender, evaluar y reformarse nuevamente. De hecho, aunque las reformas educativas de la presente década manifiestan grandes similitudes en sus componentes ideológicos y estrategias propuestas, se ejecutan y desarrollan siguiendo patrones muy diferentes en cada país. El mejor legado de tales reformas parece ser el fortalecimiento de los valores sociales en torno de la necesidad de una buena educación para todos y el establecimiento de mecanismos para su mejoramiento constante (Álvarez, 1997).

El compromiso de los gobiernos con la promoción de la competencia cognoscitiva de los ciudadanos no puede limitarse a ajustes transitorios de los sistemas educativos; este compromiso implica decisiones que afectan a largo plazo los patrones de asignación de recursos y simultáneamente, la asignación de responsabilidades. Además de resolver los inevitables dilemas sobre la naturaleza y el monto de las inversiones necesarias, es preciso definir métodos adecuados e identificar las instituciones que deben medir, valorar, desarrollar y difundir información al respecto. Aunque no sabemos mucho sobre el conocimiento como insumo y como producto en la economía, la información sobre el desarrollo, la medición y el empleo del capital humano, se está convirtiendo en una referencia esencial de las políticas nacionales. La retórica predominante atribuye a la educación un valor como inversión — en oposición a gasto— a pesar de que no disponemos de suficientes recursos teóricos y prácticos para su medida y evaluación. Tanto los economistas, como los contadores y educadores deben contribuir a la tarea de desarrollar señales, indicadores y evidencias para estimar, validar y distribuir información sobre la competencia humana, de tal manera que esta información sirva de base para acrecentar el capital social nacional para el desarrollo y la utilización del conocimiento (Organisation for Economic Co-operation and Development, 1996).

A medida que las empresas introducen mayor conocimiento e inteligencia en los procesos productivos, se incrementan sus necesidades de información para identificar y desarrollar su capital humano. Al igual que los gobiernos, las empresas y las organizaciones privadas, también las familias y los individuos necesitan decidir continuamente sobre sus inversiones en educación en relación con los beneficios económicos, personales y sociales que de ella se derivan.

Todas estas preocupaciones han incidido en una paulatina ampliación del ámbito de la evaluación del aprendizaje, que se ha extendido de un interés inicial en la selección de los individuos para cargos especiales o posteriores oportunidades educativas, al juicio sobre las organizaciones y sobre los sistemas nacionales de aprendizaje. En el Brazil, Colombia, Costa Rica, Chile, México y Venezuela, primeros países de América Latina en crear sistemas nacionales de pruebas de rendimiento escolar, la evaluación del aprendizaje escolar ha servido por décadas como mecanismo de selección para las escasas oportunidades de educación superior disponibles, en forma paralela a como sucedió en la China, donde se desarrollaron las primeras pruebas de las que tenemos conocimiento, hace cuatro milenios, en torno de las cuales se organizaba la educación del nivel superior y cuyo propósito era identificar candidatos para el servicio público en los condados, las provincias y el imperio (Wills y Lottich, 1961).[2] Pero nuevas necesidades de información están obligando a esos países a una ampliación de la gama de objetivos y enfoques de la evaluación educativa, a un mejoramiento de la competencia técnica y al desarrollo de estrategias de comunicación más efectivas.

El desafío más importante, sin embargo, no consiste en el mero desarrollo de tecnologías y sistemas de evaluación, sino en el fortalecimiento de la capacidad de la sociedad en su conjunto para intervenir activamente en proyectos de interés colectivo — como la educación— que determinan su destino; en fomentar una cultura que favorezca la evaluación, el seguimiento y la responsabilidad social y en mantener una infraestructura de conocimiento (especialistas, instituciones y redes de interacción) que permita el tránsito de sistemas cerrados y aislados de pruebas a proyectos abiertos de aprendizaje social.

DIMENSIONES Y FUNCIONES DE LA EVALUACIÓN

La evaluación busca responder a las necesidades de conocimiento práctico, de decisión y de reinvención de objetos o programas determinados. Sin embargo, la acumulación del producto de varias evaluaciones va constituyendo un soporte de aprendizaje que inevitablemente trasciende las preocupaciones inmediatas de una actividad o programa. La evaluación proporciona en efecto, un espacio para el encuentro de tres dimensiones: las utopías e ideales de una sociedad, el conocimiento disponible y la práctica social (figura 1). La evaluación no sólo facilita el acercamiento de las políticas y prácticas a las metas deseadas, sino que ayuda a trazar la línea que separa lo ideal de lo que es alcanzable.

Figura 1. Dimensiones de la evaluación

Estas tres dimensiones o perspectivas están en cierta medida presentes en todo ejercicio de evaluación, no obstante que se encuentran en permanente tensión. Algunas veces las evaluaciones están guiadas prioritariamente por la dinámica del conocimiento, otras veces por las necesidades de la práctica y, en ocasiones, por principios ideológicos, filosóficos o políticos. Pero en todos los casos, bien se trate de indicadores nacionales de desarrollo humano, de medición del éxito de las organizaciones o de la constatación del progreso personal, no es posible prescindir de los ingredientes de conocimiento, valores y utopías, que se mezclan en proporciones variadas a lo largo de un proceso de evaluación. Este proceso permite a los países, las organizaciones y los individuos compararse con otros, con sus propios ideales o con parámetros basados en la investigación científica. La evaluación de los maestros, por ejemplo, combina criterios surgidos de la práctica profesional, del estado de la investigación y de las expectativas de la sociedad a la cual sirven.

Por el hecho de encontrarse en medio de una dinámica de diversos intereses, la evaluación en educación ha servido para varios propósitos. En su versión más conocida y tradicional, los exámenes escolares, la evaluación ha tenido, entre otras, una *función simbólica*. La evaluación marca el final de un ciclo y proporciona credibilidad a los procesos educativos al acreditar frente a la sociedad el logro de determinados objetivos de aprendizaje o de formación. A una escala más amplia, la evaluación educativa ha demostrado también una *función política* puesto que está ligada a decisiones sobre el destino de las personas y a la supervivencia de las

instituciones y programas. En muchos casos la información proporcionada por la evaluación conduce al replanteamiento de políticas. Esto sucede cuando, por ejemplo, las comparaciones entre los indicadores de eficiencia o efectividad de los sistemas educativos de diversos países motivan la formulación de metas o reformas de política social en determinado país o estado.

Otras veces sucede lo inverso, es decir, la política social promueve y fomenta el desarrollo de la tecnología de la evaluación. Las recientes políticas de descentralización administrativa han motivado en varios casos el desarrollo de sistemas locales de seguimiento y supervisión de las escuelas. Este es el caso de los estados brasileros de Minas Gerais y Paraná que han desarrollado procesos de autoevaluación de las escuelas para promover su responsabilidad y su mejoramiento en un clima de mayor autonomía. En términos generales, la práctica de la evaluación es un ejercicio político, de conocimiento y de poder, así se refiera a personas, instituciones o sistemas de pequeñas localidades o de países.

La competencia internacional proporciona un impulso importante a las reformas nacionales de la educación y al desarrollo de mecanismos e instrumentos de evaluación. A finales de la década de 1950, por ejemplo, se produjo uno de los movimientos más profundos de renovacion de la educación en la historia de los Estados Unidos, como resultado de la confrontación científica con la Unión Soviética. Este movimiento introdujo cambios sustanciales en el currículo, los textos escolares, la organización de la educación y, sobre todo, ayudó a crear una corriente de pensamiento en torno de la educación que pronto se extendió a otros países, particularmente de América Latina.

Uno de los pilares conceptuales que guiaron ese proceso fue el relativo a los fines y procesos de la educación y a su relación con la medición y constatación de sus resultados. Varios de estos trabajos sobre evaluación —-que se convirtieron en clásicos de la educación— destacaron la importancia de identificar las metas que se esperaba lograr por medio de la educación y la forma como debería comprobarse el alcance de tales propósitos. Los objetivos de la educación se organizaron alrededor de taxonomías o clasificaciones que señalaban los diversos niveles de destrezas de pensamiento involucrados en el logro de diferentes objetivos.[3] Este enfoque sistemático de la enseñanza y el aprendizaje tuvo una enorme influencia en la planificación de la educación de todo el continente y en la forma de concebir sus resultados, al menos en los organismos centrales de diseño de currículo. Hasta hace poco, el currículo escolar de las escuelas paraguayas estaba diseñado con base en objetivos expresados en términos de comportamiento observable, siguiendo la tradición iniciada entonces.

Las compañías editoriales privadas contribuyeron a modernizar la enseñanza de la ciencia y la matemática y algunos gobiernos, a pesar de las limitaciones de la tecnología de evaluación disponible y la escasa demanda social existente, iniciaron o fomentaron la creación de servicios de pruebas de rendimiento escolar que aún subsisten. Esto coincidió con el desarrollo de centros de investigación públicos y

privados que se crearon a comienzos de la década de 1970 con el propósito de estudiar los problemas educativos y de promover innovaciones a lo largo de todo el hemisferio.

Paradójicamente, la conclusión de la guerra fría no disminuyó sino más bien aumentó la competencia internacional. El final del siglo es el escenario de la batalla por la conquista de los mercados. Los sistemas nacionales de educación se enfrentan nuevamente al estímulo del poder y a la comparación con sus pares. Se trata de problemas similares a los que se enfrentaron hace casi cincuenta años, pero en un contexto diferente. La competencia en una economía global basada en la innovación y el conocimiento se traduce en un mayor interés por lograr una educación para todos los ciudadanos de un país y que responda a elevados estándares de calidad. Los estándares se definen como una meta y una medida de avance en dirección a esa meta. Ravitch (1995) menciona tres tipos de estándares relacionados entre sí: estándares de contenido o de currículo (lo que debería enseñarse), de desempeño escolar (niveles de logro) y de oportunidades de aprender (disponibilidad de programas y recursos). El movimiento en favor de los estándares de la educación tiene además un horizonte internacional, particularmente en ciencias y matemáticas, que han sido el foco de la mayor parte de las comparaciones del rendimiento escolar entre países.

Este clima de competencia global hace a los sistemas educativos de los países de las Américas concientes de las tres brechas fundamentales que afrontan: la disparidad de recursos humanos existente entre los países de la región y entre las diversas regiones del mundo, las persistentes — y en algunos casos crecientes— desigualdades al interior de cada país y la gran distancia que separa el avance del conocimiento de las oportunidades de aprendizaje que estamos ofreciendo a nuestros niños y jóvenes. La evaluación permite una mejor comprensión de estas brechas y de las alternativas políticas y programáticas disponibles.

Esta consideración nos lleva a constatar que la evaluación cumple también una *función de conocimiento*. La práctica profesional de la evaluación, particularmente en los Estados Unidos, ha privilegiado sus propósitos como mecanismo sistemático de decisión sobre el valor o mérito de un objeto o programa social, con menor interés en su contribución al conocimiento sobre temas tales como la eficiencia de las intervenciones sociales y el aprendizaje de las organizaciones. El "conocimiento práctico" como lo definió Aristóteles — en oposición al "conocimiento especulativo" (el conocimiento por sí mismo)— es una categoría muy útil para describir la naturaleza del saber que genera la evaluación. Este tipo de conocimiento busca resultados útiles y se desarrolla con el fin de guiar las acciones humanas. Implica, según el filósofo, el uso de las facultades innovativas y creativas, la capacidad de organizar y el deseo de actuar. Este conocimiento estratégico sobre el funcionamiento de las intervenciones sociales y las instituciones educativas es uno de los productos más útiles de la evaluación. Sobre este conocimiento se apoya la *función de mejoramiento* de los objetos de la evaluación, que ha sido destacada como su característica primordial para distinguirla de la investigación que tendría una

finalidad de comprobación de teorías.

No menos importante que las anteriores es la *función de desarrollo de capacidad* que las instancias de evaluación pueden promover en un país. Los programas en que se basan las reformas de la educación, por ejemplo, son generalmente de intensa actividad y corta duración, en especial cuando son financiados por préstamos o donaciones internacionales. A menos que un país o estado desarrolle una infraestructura de monitoría y seguimiento que le permita repensar y revisar los cambios que ejecuta, se encontrará en una situación de perpetua incertidumbre. Esa infraestructura se compone fundamentalmente de una masa crítica de analistas, de una base institucional y de redes de interacción e información. Es decir, de recursos humanos capacitados para las tareas de evaluación, organizaciones públicas o privadas responsables y eficientes canales de comunicación y participación.

EVOLUCIÓN DE LA PRÁCTICA DE LA EVALUACIÓN EDUCATIVA EN EL HEMISFERIO

Dos vertientes principales confluyeron en el desarrollo de la evaluación educativa durante el siglo XX en el hemisferio americano. La primera, se originó en el seno de la psicología, como parte de la utilización de las teorías e instrumentos de medición elaborados en torno de los conceptos de inteligencia, aptitud, rendimiento y aprendizaje escolar. La segunda, surgió de la progresiva adaptación de las ciencias sociales a las crecientes necesidades de orientación y guía por parte de los programas de intervención, las políticas y las inversiones sociales.

La tecnología y la práctica de la evaluación de diversas características individuales y del aprendizaje se han enriquecido notablemente desde que se emplearon las primeras pruebas o "tests" en Francia —desarrolladas por Binet— con el propósito de identificar a los niños retardados. Buyse, discípulo de Binet, y otros profesores europeos introdujeron las ideas fundamentales de la naciente psicometría en Colombia durante la década de 1930 (Restrepo, 1995). Desde entonces, Colombia, junto con otros países como el Brazil y Chile se convirtió en pionera en este campo en América Latina. Pero los Estados Unidos ha sido el país donde la demanda por los servicios de pruebas ha sido más intensa y donde la oferta se ha multiplicado más ampliamente.

En la Universidad de Costa Rica se organizó un servicio de pruebas en 1986 con la finalidad de mejorar el sistema educativo, más que de ayudar a las universidades en sus procedimientos de admisión de estudiantes, lo cual constituía el objetivo predominante de los servicios de pruebas existentes entonces (Esquivel, 1996). Lo mismo hizo Chile a través de su Sistema de Medición de la Calidad de la Educación en 1988. Varios países de la región han optado por un esquema similar, de tal manera que los nuevos servicios de pruebas desarrollados en la presente década se orientan a mejorar la calidad del sistema educativo en contraposición a sus antecesores. El cuadro 1 presenta información sobre los países que desarrollaron inicialmente procesos selectivos de medición del rendimiento escolar y una segunda

generación de sistemas (casi todos ellos posteriores a 1985) inspirados en la analogía de la empresa eficiente, que intentan servir de referencia y estímulo al mejoramiento de la calidad de la educación. Está comenzando a emerger una tercera generación de enfoques de evaluación de carácter más amplio que utiliza analogías múltiples y busca afectar el sistema educativo en diversas áreas críticas, además de sus resultados finales en términos del rendimiento escolar.

La evaluación educativa ha adquirido, además, carta de ciudadanía en varios países. Se encuentra mencionada en la Ley General de Educación de 1994 de Colombia (Restrepo, 1995), en una Ley Orgánica Constitucional de Chile promulgada durante el régimen militar (Rodríguez, 1996) y en la Ley de Directrices y Bases de la Educación Nacional del Brasil (1996), la cual incluye entre las tareas de los organismos nacionales de educación: "Asegurar un proceso nacional de evaluación del rendimiento escolar en la enseñanza fundamental, media y superior en colaboración con los sistemas de enseñanza, estableciendo una definición de prioridades y una mejora de la calidad de la enseñanza."

En algunas de las constituciones nacionales, se comienzan a percibir los efectos del cambio en el papel del Estado en relación con la educación y el nuevo énfasis en garantizar no solamente el acceso a la educación, sino también su calidad. La Constitución Nacional del Perú, en el Artículo 16 establece que:

> Tanto el sistema como el régimen educativo son descentralizados. El Estado coordina la política educativa. Formula los lineamientos generales de los planes de estudio así como los requisitos mínimos de la organización de los centros educativos. Supervisa su cumplimiento y la calidad de la educación.

También las leyes estatales de educación en los países más grandes comienzan a reflejar la preocupación por asegurar que la actividad educativa produzca los resultados esperados, como lo ilustra la reciente Ley de Educación del Estado de Aguascalientes en México (1997), que autoriza al Instituto de Educación a crear un Sistema Estatal de Evaluación Educativa.

Aunque la investigación sobre la educación floreció en todo el continente con la vertiginosa expansión de los sistemas educativos de las décadas de 1960 y 1970, el desarrollo de una metodología y un abordaje sistemático de la evaluación de los programas y las instituciones se inició apenas recientemente en la mayoría de los países del hemisferio con excepción del Canadá y los Estados Unidos. En este último país, el impulso a los grandes programas de intervención social en favor de la infancia y juventud en desventaja social, promovidos por el gobierno federal hacia la mitad de la década de 1960 por medio de donaciones que requerían de evidencias empíricas sobre su efecto, fortaleció el desarrollo de un movimiento en torno de la evaluación que condujo a la creación y experimentación de modelos de trabajo y configuración de una comunidad académica (Worthen y Sanders, 1987). Este movimiento fue precedido, sin embargo, por una primera etapa de creación de las bases conceptuales en las décadas de 1940 y 1950.

Cuadro 1. Sistemas nacionales de pruebas de rendimiento escolar

Fuentes: Banco Interamericano de Desarrollo (1997); Álvarez (1995)

Esta comunidad ha buscado identificar principios que orienten la práctica profesional de la evaluación, tales como los estándares para el desarrollo y la aplicación de pruebas psicológicas y educativas en 1966, los estándares para la evaluación de programas educativos (The Joint Committee on Standards for Educational Evaluation, 1994) y los estándares para la evaluación de personal en 1988.

En América Latina la investigación educativa ha sido muy rica en sus concepciones teóricas y en el diseño de innovaciones estrechamente relacionadas con la problemática social de la región, particularmente las brechas de desarrollo y la alienación de las grandes mayorías. Pero ha sido menos productiva en materia de medición y análisis de los resultados y de la calidad de la educación, las cuales constituyen el propósito de la mayoría de los programas de revitalización de la educación que se están ejecutando en casi todos los países de la región con apoyo de los bancos y las agencias internacionales.

La información recogida en 1996 sobre una muestra de 60 programas educativos en operación o en proceso de aprobación, financiados por el Banco Mundial, el Banco Interamericano de Desarrollo (BID) y la Agencia de los Estados Unidos para el Desarrollo Internacional (USAID) (Academy for Educational Development, 1996), muestra que una gran mayoría de tales programas (67 por ciento) poseen un componente de evaluación. En el 13 por ciento de los casos se trata explícitamente de exámenes o pruebas de rendimiento escolar (figura 2). Sin embargo, este renovado interés por conocer los efectos del sistema escolar contrasta con las

notables deficiencias de los sistemas de información sobre educación en América Latina y el Caribe. Este es uno de los problemas que ha encontrado la Secretaría protémpore de las Américas en su esfuerzo colectivo de seguimiento de las metas de desarrollo humano en el hemisferio.

Figura 2. Evaluación educativa en los proyectos internacionales

Fuente: Academia para el Desarrollo Educativo (1996)

No obstante el crecimiento y expansión de las actividades de evaluación en la educación en el continente americano, existen todavía para la comunidad científica y el público interesado en la educación grandes limitaciones para influir en las decisiones políticas. No se ha desarrollado suficientemente una "cultura" de la evaluación dentro de la cual tengan sentido sus diferentes componentes y que permita una concertación de los diversos grupos que componen la comunidad educativa nacional. No obstante, se cuenta con un acopio de experiencias, conocimientos e instrumentos que pueden prestar una ayuda de primer orden a los países y sistemas escolares para la asignación de recursos y el mejoramiento de las oportunidades de aprendizaje y desarrollo humanos.

ÁREAS CRÍTICAS DE LA EVALUACIÓN DE LA EDUCACIÓN

La fragilidad que caracteriza la relación entre la comunidad científica, las familias y escuelas y los organismos de decisión es particularmente preocupante en América Latina y el Caribe por dos motivos principales. En primer lugar porque, a pesar de la retórica política predominante, no se ha logrado establecer una actitud de responsabilidad colectiva en torno de la educación pública ni se han establecido mecanismos eficientes de monitoría del rendimiento de los sistemas escolares. La

educación como proyecto de la sociedad en su conjunto es todavía una utopía lejana. Se podría extender a casi todos los países de la región el argumento de que el obstáculo más grande para la reforma educativa del Brazil es el predominio del interés privado sobre las finalidades públicas (Plank y otros, 1994). En segundo término, porque las nuevas políticas sociales tienden a asignar responsabilidades nuevas a comunidades sin recursos y con escasa capacidad de convocatoria.

A la anterior debilidad de la capacidad de la sociedad civil para intervenir en las decisiones educativas nacionales, se suman las limitaciones de los mismos sistemas para controlar la empresa educativa. Un reciente análisis realizado por el Banco Interamericano de Desarrollo (1997) de los estudios sectoriales sobre la educación primaria realizados en 15 países de América Latina y el Caribe entre 1991 y 1995 y financiados por el Banco Mundial y el BID, mostró que en el 53 por ciento de tales países había una carencia de evaluación de resultados en términos de aprendizaje de los estudiantes; en el 40 por ciento no existían sistemas de información operativos; en el 67 por ciento había un sistema inadecuado de supervisión de los maestros y en el 40 por ciento existía una limitada capacidad del Ministerio de Educación para planear, investigar y evaluar (Cuadro 2). Esto conduce a pensar que la tarea es doble. Por una parte, fortalecer la capacidad nacional y, por otra, atender a las necesidades urgentes de evaluación de los programas de reforma en operación.

Hay cuatro áreas de información y evaluación que son críticas para la política social y que en materia de educación un país o estado tiene que definir. Estas áreas están estrechamente relacionadas entre sí.

- La primera es la relativa a los resultados de la suma de los esfuerzos y empresas educativas de la sociedad.
- La segunda es la relevancia y eficiencia del proceso de la educación.
- La tercera área crítica es el profesionalismo de los actores que lideran el proceso: maestros y administradores.
- La cuarta es el contexto inmediato de aprendizaje, que en la mayoría de los casos se trata de la escuela.

Seguimiento de los resultados de la educación en el ámbito nacional

La educación formal ha contado siempre con una forma de acreditación de sus efectos a escala individual por medio de exámenes, calificaciones y diplomas. Pero a medida que se ha ampliado la historia educativa de las poblaciones, más personas estudian durante períodos más largos y sobre temas cada vez más vastos, se ha incrementado la necesidad de información agregada sobre la educación a niveles estatales, nacionales y globales. La información proveniente del sector educativo es a su vez un ingrediente para el análisis del progreso de los países y para el diseño de políticas sociales. Aunque desde la época de la rápida expansión de los sistemas educativos de América Latina y el Caribe (décadas de 1960 y 1970) se comenzó a recolectar información educativa, existe un acuerdo generalizado en que ésta ya no

es suficiente como insumo para conocer los resultados del esfuerzo educativo nacional, ni para emprender acciones para mejorarlos.

Cuadro 2. Problemas de evaluación identificados en proyectos internacionales en 15 países de América Latina, 1991-1995

Fuente: Banco Interamericano de Desarrollo (1997)

La información disponible sobre los resultados de la educación consiste, en la mayoría de los casos, en estadísticas sobre indicadores de aspectos formales de los sistemas (acceso, retención, etc.) de dudosa confiabilidad y algunos datos obtenidos en algunos países a través de pruebas y exámenes. Los países del Caribe de habla inglesa, siguiendo la tradición europea, desarrollaron un sistema de exámenes públicos nacionales mientras que algunos de los países de Centro y Sur América (Cuadro 1) cuentan con información proveniente de la aplicación de pruebas estandarizadas o de algún tipo de mediciones del rendimiento escolar basadas en muestras de estudiantes de educación básica. No obstante, estos recursos son insuficientes para responder a las necesidades creadas por el nuevo énfasis en la calidad de los resultados de los sistemas educativos nacionales. A este respecto, la política educativa de finales del siglo en todos los países de las Américas se enfrenta a interrogantes como los siguientes:

- ¿Cuáles son los propósitos, la clientela y la utilidad de los sistemas nacionales de evaluación?
- ¿Qué información deberían incluir los sistemas nacionales de evaluación del aprendizaje?
- ¿Es el establecimiento de metas y la evaluación de los resultados de la educación un asunto de política nacional o de decisión local?
- ¿Cuál es la contribución de las metas y estándares al mejoramiento de la educación?
- ¿Es deseable, conveniente y factible la formulación de estándares internacionales para el aprendizaje escolar?

Los dilemas que tienen que resolver los países con escasos recursos son aún más apremiantes. Un asunto fundamental para ellos es estimar si los beneficios de la inversión en los sistemas de seguimiento de los resultados de la educación supera a los que se lograrían por medio del mejoramiento de los insumos o de la organización de la educación.

Evaluación del proceso de escolarización

El conocimiento sobre los resultados de la educación es un elemento esencial para la evaluación de los sistemas educativos, pero no es suficiente. La aplicación de pruebas, mediciones o constataciones del rendimiento escolar no eleva por sí misma la calidad del servicio. Sin una información constante sobre lo que se enseña y cómo se enseña, no es posible explicar los resultados ni proporcionar orientaciones sobre las prácticas deseables y los requisitos profesionales de los docentes.

En un clima de reforma y formulaciones grandiosas como el presente, es oportuno recordar las lecciones de la historia. Buena parte de las transformaciones de currículo del pasado no lograron establecerse plenamente porque no pudieron afectar la vida cotidiana del microcosmos de la escuela y el aula. Las decisiones en torno de la política de seguimiento del currículo escolar no están exentas de tensiones y dilemas. En efecto, si se acepta que las metas de la educación deben ser formuladas por el gobierno nacional con el objeto de obtener resultados de alta calidad, ¿qué espacio de participación en el corazón del proceso educativo queda para las comunidades escolares autómomas, los profesionales de la educación y las múltiples opciones pensables en una sociedad democrática?

- ¿Qué tipo de monitoría permitiría un balance entre el currículo nacional deseable, el planeado en cada localidad y escuela y el real? ¿Se trata de la relación existente entre el qué enseñar determinado centralmente y el cómo enseñar realizado localmente?
- ¿Qué políticas asegurarían una coherencia entre el currículo escolar, los sistemas de pruebas y exámenes y los textos y materiales escolares?
- ¿Deben las reformas educativas orientarse por la innovación del currículo, por la

reestructuración de la organización escolar o por los estándares y evaluación de resultados?

Evaluación de la organización escolar

La preocupación por la eficiencia de las organizaciones escolares, ha estado sin duda motivada por lo que sucede en el campo de la organización productiva contemporánea. Pero tal preocupación no es totalmente nueva en el Continente. Hace un poco más de un siglo, la Corona Británica introdujo, primero en Barbados y posteriormente en otros países del Caribe, un sistema de pago por resultados para las escuelas que estaba basado en exámenes diseñados "para probar el carácter de la enseñanza impartida y la administración general de la escuela primaria" (King, 1995). Los inspectores de entonces preparaban "estándares" y puntajes y establecían rangos de eficiencia a ser utilizados como insumo para establecer políticas de financiación. Los interrogantes que nos planteamos en la actualidad no son quizás muy diferentes. Por ejemplo:

- ¿Es justo evaluar una escuela por los resultados de sus estudiantes en comparación con otros estudiantes de otras escuelas?
- ¿Qué tanto puede la escuela aprender de las organizaciones productivas en su manejo y evaluación, sin perder su carácter y olvidar su finalidad?
- ¿Cómo se definiría una escuela eficaz?
- ¿Qué condiciones y factores organizativos sugiere la evidencia empírica estar asociados al éxito de la escuela?

Evaluación en la profesión de maestro

Aunque el oficio de maestro no tenga el mismo halo de dignidad que lo rodeó a lo largo de la historia, su éxito, medido en términos de su efecto en la vida de los estudiantes, es quizás más necesario ahora que en el pasado para el progreso de los pueblos. Pero la importancia cualitativa y cuantitativa de la profesión docente contrasta con el poco interés demostrado por los gobiernos, la sociedad en su conjunto y los mismos maestros en la profesionalización de sus actividades. Profesionalización que conlleva responsabilidades sociales y obviamente mecanismos de selección, promoción e incentivación dentro de la carrera docente, así como de evaluación de su desempeño.

Problemas de conceptualización sobre cuáles son los ingredientes fundamentales de una enseñanza efectiva y el escaso interés de los investigadores, problemas políticos tales como la prevalencia de ministerios de educación débiles frente a sindicatos fuertes y problemas administrativos tales como la centralización e inexistencia de incentivos para los maestros, han impedido el desarrollo de mecanismos para mejorar la competencia profesional de los maestros. La evaluación es, en efecto, útil no solamente para los administradores y padres de familia sino también para el maestro mismo. No obstante, en este terreno, la política enfrenta interrogantes y

dilemas de difícil solución, entre los cuales se pueden mencionar los siguientes:

- ¿Cuál es el propósito de la evaluación de los maestros?
- ¿Cuáles son los criterios para evaluar el desempeño de los docentes?
- ¿Quién es el responsable de la evaluación de los maestros?
- ¿En qué períodos de la carrera profesional de los maestros es más útil la evaluación?
- ¿Qué nos enseña la experiencia de los ensayos de evaluación existentes?

Este libro aborda las cuatro áreas críticas de evaluación y política social mencionadas anteriormente, es decir: los resultados de la educación formal, los procesos de la educación, los maestros y las organizaciones escolares, combinando la información proveniente de tres fuentes: la investigación internacional, la experiencia de los países con mayor trayectoria en evaluación, y la ejecución de recientes innovaciones que, en materia de evaluación han sido motivadas por las reformas educativas de la década de 1990 en varios países del hemisferio. Su propósito es contribuir al aprendizaje de las naciones por medio de la identificación y el análisis de los dilemas y opciones que se plantean individual y colectivamente en los países, los estados y las familias.

El libro está organizado en tres partes. La primera parte, "los sistemas nacionales"se inicia con un capítulo sobre la monitoría de los resultados de la educación en el ámbito nacional, el cual proporciona un escenario global al tema del seguimiento del desempeño de los sistemas nacionales de educación. Para la mayoría de los países, especialmente los de recursos más escasos, el dilema consiste en decidir invertir en la evaluación del rendimiento escolar o en los insumos básicos para el funcionamiento del sistema. La primera opción conduce a nuevas encrucijadas. ¿Qué tipos de seguimiento de los efectos de la educación son oportunos y utilizables?, ¿qué recursos institucionales, técnicos y financieros se requieren? Aunque el interés en el tema de los resultados agregados de la educación es relativamente nuevo, varios países poseen una experiencia considerable en la evaluación del aprendizaje escolar siguiendo diversos esquemas, los cuales han servido de referencia para la consideración de los productos globales de los sistemas educativos. Este capítulo integra la información aportada por la experiencia internacional con el análisis de los resultados de la investigación social y con los problemas prácticos que la decisión política encuentra en el desarrollo de los sistemas nacionales de seguimiento del aprendizaje escolar.

El siguiente capítulo presenta un análisis de la evaluación de los procesos que supuestamente conducen a los resultados alcanzados por los sistemas educativos. Su eje temático es el seguimiento y la monitoría del currículo escolar. El currículo escolar es, en efecto, un instrumento central de la política educativa. El capítulo ilustra la relación entre evaluación y política en el contexto de recientes análisis y comparaciones internacionales del aprendizaje escolar. La primera parte del libro concluye con la discusión del caso de monitoría internacional de las metas de

desarrollo humano que los países de las Américas iniciaron como resultado del Acuerdo de Nariño firmado en 1994.

La segunda parte, "las lecciones de la historia," intenta recuperar la experiencia de dos de los países de América Latina con mayor tradición en evaluación, por medio de la identificación y discusión de los efectos sociales que los sistemas de evaluación educativa han tenido a lo largo de las últimas décadas. La cuestión fundamental que estos capítulos discuten es qué ha significado para las familias y los estudiantes, para la política social y para la organización de la educación en los respectivos países, la operación de un sistema de evaluación periódica del rendimiento escolar y qué lecciones se pueden derivar de esta experiencia para la política educativa del futuro.

La tercera parte presenta un estado del arte sobre la temática de la evaluación de los maestros, complementado con una revisión de la práctica de la evaluación de maestros en cinco países de América Latina. Además de las dificultades teóricas mencionadas anteriormente, la evaluación de los maestros está sometida a presiones políticas de los sindicatos de educadores y de grupos y asociaciones profesionales y se encuentra limitada en muchos países por una indefinición de sus propósitos. A esto se añade el hecho de que la nueva dinámica creada por los procesos de descentralización y autonomía escolar está obligando a los gobiernos locales y a las comunidades escolares a buscar mecanismos de control y desarrollo de la profesión docente. Los gobiernos de la región se encuentran frente a un dilema paralelo al que encontró Plinio el Joven cuando quizo establecer una escuela en su pueblo y decidió financiar solamente una parte de su costo con el siguiente razonamiento:

> Si financio todos los costos, mi donación puede ser utilizada para propósitos egoístas, como sucede en muchos lugares en que los salarios de los maestros son sufragados por fondos públicos. Solamente hay un remedio para enfrentar ese mal; que el nombramiento de maestros se deje completamente en las manos de los padres de familia y que éstos tengan una obligación de contribuir a los gastos (Carta XIII).[4]

El problema de la financiación de los sistemas educativos, la responsabilidad social de sus directivos, el desarrollo profesional de los maestros y su influencia en la calidad de los resultados de la educación, están unidos por las instancias de evaluación disponibles, las cuales son en general, tímidas, insuficientes e inoperantes.

La tarea fundamental de la educación es proporcionar un ambiente adecuado y un estímulo para el aprendizaje. La escuela representa el prototipo de este contexto. Allí confluyen las expectativas, los actores y los recursos para el aprendizaje. El capítulo con el que se inicia la cuarta parte de este documento proporciona un marco de referencia para ubicar las opciones de política encaminadas a evaluar las escuelas dentro del contexto de reformas nacionales o estatales. Inscribe tales opciones en la tradición académica de la evaluación en general y de la investigación sobre la efectividad de la escuela en particular y presenta un modelo alternativo, el

del valor agregado, que ha sido experimentado en el Estado de Texas en los Estados Unidos. Este modelo incluye no solamente la evaluación de las escuelas en particular, sino también la de los sistemas escolares en general. El problema de la eficiencia de la empresa educativa en América Latina ha sido destacado por recientes análisis y comparaciones internacionales (Banco Interamericano para el Desarrollo, 1996).

Los capítulos que le siguen describen dos de las innovaciones que varios países están ensayando (el estado de Paraná en el Brasil y el estado de Aguascalientes en México) con el ánimo de acompañar y evaluar el proceso de reforma de la educación en el enclave mismo en que se desarrolla. El documento concluye con un breve análisis de los dilemas que emergen y las alternativas de política por las que los países pueden optar.

CRITERIOS PARA UNA AGENDA DE POLÍTICAS DE EVALUACIÓN

La evaluación de la educación es un componente esencial del mecanismo de aprendizaje de las naciones. Ésta permite no solamente conocer los avances y rezagos de las oportunidades y logros de sus gentes, y emitir juicios más certeros, sino también avanzar en la búsqueda de ideales colectivos. Solamente por medio de la reflexión provocada por la evaluación, se puede trascender la retórica predominante que, en un clima de reforma como el presente, puede impedir la exploración de nuevas alternativas. Como instrumento de aprendizaje social, debe cumplir requisitos provenientes de tres tipos de criterios:

- Criterios de relevancia y utilidad social
- Criterios de orientación al futuro y
- Criterios de precisión.

En efecto, la evaluación no es un asunto de carácter exclusivamente técnico puesto que sin una base social amplia y una capacidad e interés de la comunidad nacional que se traduzca en una voluntad política, las instancias de evaluación son inoperantes y socialmente inexistentes. Más aún, las actividades de concertación y negociación implicadas en el ejercicio de la evaluación constituyen uno de los canales más útiles de interacción de una sociedad en torno de la construcción de su propia identidad. Por tanto, una política nacional de evaluación debe responder a las necesidades de integración y participación de los diversos grupos de la sociedad con responsabilidad e interés en la educación. Así mismo, requiere de una visión comprensiva de las áreas que constituyen los temas críticos de decisión.
Aunque las evaluaciones tienen como objeto inmediato las huellas del pasado y la información del presente, cobran sentido y contribuyen al desarrollo humano y social en la medida en que su propósito es una determinada acción sobre el futuro, que es su campo de aplicación y existencia. El segundo grupo de criterios para una agenda de política de evaluación, comprende entonces, aquéllos criterios que ayudan a que las actividades de evaluación sirvan al futuro de los sistemas

educativos y al fomento de su capacidad de cambio.

La calidad técnica de la evaluación es un requisito indispensable para su confiabilidad y uso. Esta calidad no se limita sino que se amplía por la incorporación de metodologías alternativas y de disciplinas y metáforas nuevas. Evaluar en las lenguas romances tiene el significado original de pesar y evoca la imagen de la balanza que permite emitir juicios con conocimiento y precisión. El símbolo de la precisión era, para los antiguos egipcios, la pluma (la diosa Mayet, representante de la justicia, la verdad y el orden del mundo) que servía de pesa en el platillo de la balanza en que se pesaban las almas.[5] La precisión es, en efecto, la característica que permite un juicio ponderado y genera un conocimiento práctico sobre los que se construye el aprendizaje de las naciones.

NOTAS

[1] T. de Chardin, por ejemplo, había previsto una progresiva humanización y expansión de la red de pensamiento en simultánea diversificación y convergencia.

[2] El complejo sistema de pruebas en la China servía de filtro de seleccción desde los condados hasta la capital. Los candidatos que lograban pasar las pruebas nacionales tenían derecho a ocupar puestos públicos (Wills, E. y Lottich V. 1961. *The Foundations of Modern Education*. New York: Holt, Rinehart and Winston)

[3] La taxonomía de los objetivos de la educación propuesta por una serie de expertos encabezados por Benjamin Bloom de la Universidad de Chicago en 1956, representa muy bien este movimiento.

[4] Citado por Coulson, A. (1996). "Markets versus Monopolies in Education: the Historical Evidence." *Education Policy Analysis Archives*, 4, 9.

[5] Citado por Calvino, I. (1994). *Seis propuestas para el próximo milenio*. Madrid: Ediciones Siruela.

REFERENCIAS

Academy for Educational Development. (1996). *Educational Reform Projects Supported by the World Bank, the Inter-American Development Bank, and USAID in Latin American and the Caribbean*. AED: Washington, DC. Fotocopiado.

Álvarez, B. (1995). "Equity and Selective Access to Education in Latin America". En: T. Kellaghan (ed). *Admissions to Higher Education: Issues and Practice*. Princeton: International Association for Educational Assessment.

Álvarez, B. (1997). "Life Cycle and Legacy of Educational Reforms in Latin America and the Caribbean." Presentado en la reunión anual de la Comparative International Education Society, México, marzo 19-23.

Banco Interamericano de Desarrollo (1997). *Progreso Económico y Social en América Latina. Informe 1996*. Washington, D.C.

Esquivel, J. (1996). "Medición de logros del aprendizaje y el empleo de los resultados en Costa Rica." Presentado en el seminario internacional de evaluación y estándares de la educación en América Latina. Realidades y desafíos. PREAL, Río de Janeiro, 4-5 de diciembre.

King, R. (1995). "Education in the British Caribbean: The Legacy of the 19th Century." *LaEducación. Revista interamericana de desarrollo educativo*, 34,121: 243-260.

Organisation for Economic Co-operation and Development (1996). *Measuring What People Know: Human Capital Accounting for the Knowledge Economy*. París.

Plank, D., Sobrinho, A. y Xavier, A. (1994). "Obstacles to Educational Reform in Brazil." *La Educación. Revista Interamericana de desarrollo educativo*, 33, 117: 75-95.

Porter, M. (1990). *The Competitive Advantage of Nations*. Boston: Harvard University Press.

Ravitch, D. (1995). *National Standards in American Education: A Citizen's Guide*. Washington, DC: Brookings Institution.

Reich, R. (1990). *The Work of Nations: Preparing Ourselves for the 21st Century Capitalism*. Nueva York: Vintage Books.

Restrepo, G. (1995). "Exámenes nacionales universitarios de ingreso y de egreso. Su relación con el sistema nacional de la calidad educativa". En *Misión nacional para la modernización de la universidad pública*. Bogotá: Editorial Presencia.

Rodríguez, C. (1996). "Sistema de medición de la calidad de la educación. Características y usos de los resultados de la evaluación para mejorar la calidad." Presentado en el seminario internacional sobre la evaluación de la educación en América Latina. Realidades y desafíos. PREAL, Río de Janeiro, 4-5 de diciembre.

Romer, P. (1993). "Two Strategies for Economic Development. Using Ideas and Producing Ideas." Procedings of the World Bank Conference on Development Economics.

The Economist. (1997). *Education and the Wealth of Nations*, marzo 29.

The Joint Committee pon Standards for Educational Evaluation. (1994). *The Program Evaluation Standards: How to Assess Evaluations of Educational Programs*. 2nd. edition, Londres: Sage Publications.

Wills, E. Y Lottich, V. (1961). *The Foundations of Modern Education*. Nueva York: Holt, Rinehart and Winston.

Worthen, B. Y Sanders, J. (1987). *Educational Evaluation: Alternative Approaches and Practical Guidelines*. Nueva York: Longman.

Parte I
SISTEMAS NACIONALES

CAPÍTULO 2

SEGUIMIENTO DE LOS RESULTADOS EDUCATIVOS NACIONALES

Thomas Kellaghan

La presión por lograr resultados de los sistemas educativos que se equiparen con las expectativas de las naciones ha aumentado considerablemente, con la introducción de las reformas de la política social con las que se inaugura el siglo XXI. Tanto los políticos como los académicos y los padres de familia a quienes se están abriendo canales de participación en la educación, requieren información sobre el desempeño no sólo de los estudiantes como individuos, sino también de los sistemas educativos en su conjunto. Los dilemas de política son inevitables cuando es necesario definir dónde concentrar los esfuerzos y qué fuerzas han de guiar los cambios. Este capítulo responde al primer grupo de preocupaciones planteadas en la introducción, es decir, las relativas a los resultados de la suma de los esfuerzos y empresas educativas de la sociedad, por medio de un interrogante sencillo, pero de gran alcance: ¿qué aprenden los estudiantes en la escuela? Esta pregunta permite abordar la problemática de la evaluación de los sistemas nacionales de educación desde múltiples perspectivas: la del investigador, la del administrador y la del político. Y permite también organizar los elementos del debate: la experiencia internacional, la investigación reciente y las opciones de política, en torno del escenario y de los procedimientos involucrados en el montaje y operación de un sistema nacional de evaluación.

INTRODUCCIÓN

Durante mucho tiempo la mayoría de los departamentos gubernamentales han recogido y publicado rutinariamente estadísticas que reflejan el funcionamiento y desarrollo de sus sistemas educativos. Se proporcionan estadísticas sobre diversos insumos—cifras escolares, instalaciones y matrículas—e indicadores de eficiencia tales como la proporción de alumnos por maestro y las tasas de repetición,

abandono y terminación por determinadas cohortes. Sin embargo, y a pesar de un obvio interés por conocer los logros educativos y las grandes inversiones humanas y financieras involucradas, hasta muy recientemente pocos países industrializados o en vías de desarrollo recopilaban y distribuían información sobre los resultados de la educación de una forma sistemática.

Tradicionalmente se ha asociado la información sobre los resultados—lo que los niños aprenden en la escuela—con la evaluación de individuos y se ha utilizado con fines diversos, incluidos el seguimiento del progreso de los alumnos, el diagnóstico de áreas problemáticas de aprendizaje, el incentivo de los alumnos y la orientación para promover acciones correctivas. Todos estos usos se pueden calificar de formativos, es decir, que la información derivada de la evaluación se utiliza para intervenir en el aprendizaje de los alumnos con el fin de mejorarlo. La evaluación puede ser también sumativa, cuando proporciona información sobre los conocimientos y las destrezas de los alumnos al finalizar un programa educativo.

En los últimos años, la evaluación se ha extendido a los sistemas escolares, donde se ha utilizado con fines análogos a aquéllos utilizados en el caso de los estudiantes individualmente considerados. También cuando nos referimos a sistemas, la evaluación puede ser sumativa, al presentar los resultados sin hacer referencia al tratamiento, o formativa, en cuyo caso se utiliza para el diagnóstico, el seguimiento, y la orientación de acciones correctivas. Cuando la evaluación se utiliza a nivel del sistema, podemos hablar de evaluación nacional, tal como se denomina en los Estados Unidos, si bien este tipo de evaluación no es exclusiva de este país. No obstante, los Estados Unidos aportan un modelo ejemplar de sistema de evaluación nacional—la Evaluación Nacional del Progreso Educativo (National Assessment of Educational Progress, NAEP)—que llevó a cabo su primera prueba en 1969. En la Gran Bretaña, la Unidad de la Evaluación del Rendimiento (Assessment of Performance Unit) entró en funcionamiento en 1975. Desde entonces, el sistema británico ha sido sustituido por uno más complejo que ha generado gran controversia. También hay un sistema de evaluación nacional en Francia y en muchos otros países del mundo: en varios países europeos, en el Canada y en los Estados Unidos, en América Latina (Costa Rica, Chile, Colombia, Venezuela), en África (por ejemplo, Egipto, Mali, Mauricio y Marruecos) y en Asia (por ejemplo, la China, Hong Kong y Tailandia)[1].

La problemática de los resultados de la escolarización se resume en una pregunta relativamente simple, a saber, ¿qué están aprendiendo los niños en la escuela? La respuesta, sin embargo, es bastante compleja y plantea numerosos problemas. En primer lugar, dado que la educación persigue una serie de metas, ¿qué resultados se deben tener en cuenta como representativos de los efectos de la enseñanza?, ¿quién debe escogerlos? y ¿de qué manera?. Segundo, ¿qué métodos se pueden utilizar para medir los resultados del sistema educativo?. Tercero, una vez que se ha decidido sobre el método de evaluación, ¿cómo podemos estar seguros de que la información que éste nos proporciona representa realmente los logros del sistema educativo (o de parte del mismo?). En cuarto lugar, ¿cúal es la mejor manera de presentar los

resultados? Finalmente, si deseamos iniciar una acción formativa, ¿cómo podemos transformar en acción los resultados sobre los éxitos del sistema educativo?

El fin del presente estudio es el de responder a estas preguntas, que son relevantes para los directivos de la educación, quienes se pueden estar planteando la introducción de un sistema para evaluar y supervisar los resultados de su sistema educativo. Se tendrán en cuenta la naturaleza y el fundamento de las evaluaciones nacionales, así como sus múltiples componentes, formas que adoptan en diferentes países y costo.

FUNDAMENTO DE LAS EVALUACIONES NACIONALES

Formalmente, se puede definir la "evaluación nacional" como un procedimiento concebido para medir los resultados de aprendizaje de un sistema educativo. Más adelante discutiremos los fines y usos de una evaluación nacional; por ahora, diremos que la finalidad general de una evaluación nacional es la de proporcionar información sobre los logros de los alumnos a quienes participan en la formulación de políticas y la toma de decisiones en educación con el fin de mejorar sus decisiones y sugerir intervenciones efectivas en términos de costo para mejorar el aprendizaje.

Muchos factores han llevado a los países a interesarse por los resultados del aprendizaje y a prepararse para invertir recursos con el fin de averiguar qué conocimientos, destrezas y aptitudes están adquiriendo los alumnos en los diferentes niveles del sistema educativo. Es importante indicar que ya no es admisible considerar los insumos como sustitutos de los resultados. Es decir, que no podemos suponer que porque un niño haya ido a la escuela durante cuatro, cinco o seis años, éste haya realmente adquirido, en palabras de la Declaración Mundial de Educación para Todos, "conocimientos útiles, aptitud para razonar, destrezas y valores" (UNESCO, 1990). Una segunda razón para establecer sistemas de evaluación nacional se debe al hecho de que los cambios económicos y tecnológicos exigen a los graduados escolares niveles de conocimientos y destrezas cada vez más altos y se necesita algún tipo de seguimiento para informar a los responsables de las políticas y directivos sobre los logros obtenidos en esa dirección.

La necesidad de mejorar el rendimiento académico de los alumnos también se ve acentuada como consecuencia del aumento del libre comercio y la competencia económica entre las naciones. Es en este contexto que ha retomado vigor la confianza en el papel que juega la educación en el mejoramiento del "capital humano" de un país. Hay quienes defienden que si un país no tiene un sistema educativo eficiente, tampoco contará con la mano de obra, ni con la preparación, el rendimiento y la competitividad necesarios para mantener y mejorar su desempeño económico y aumentar el nivel de vida (Guthrie, 1991). Aún más, se debe hacer un seguimiento sistemático de los resultados del sistema educativo de tal manera que se pueda asegurar a los gobiernos que dichos resultados son satisfactorios.

Una razón más para el establecimiento de evaluaciones nacionales es que éstas ayudan a responder a la cuestión de la necesaria mejora de la eficiencia del sistema educativo. Esto se debe a que en muchos países en la actualidad, los gobiernos se enfrentan con el problema de hacer frente a un aumento de la tasa de matrícula, al mismo tiempo que tratan de mejorar la calidad de la educación sin incrementar los gastos. Esto exige ampliar la eficiencia de los sistemas educativos. No obstante, para poder probar que se ha logrado este fin, es necesario tener información no sólo sobre los insumos, sino también sobre los resultados (Lockheed y Hanushek, 1988).

Si bien puede ser difícil para un país, tanto industrializado como en vías de desarrollo, hacer frente a la presión que se deriva de estas consideraciones, los países menos desarrollados deberían pararse a pensar seriamente si les conviene o no iniciar una evaluación nacional. La pregunta que deben plantearse es la siguiente: ¿Debe un país usar sus escasos recursos en emular la práctica del seguimiento de los resultados educativos de los países más ricos o convendría más que utilizara estos recursos para mejorar los insumos educativos, tales como los edificios escolares, la formación del profesorado, los libros de texto, los laboratorios u otras instalaciones?

INDICADORES

El tipo de tarea que se realiza para llevar a cabo una evaluación nacional no suele diferir mucho de la que acometería un alumno para evaluar su propio progreso escolar. Por tanto, si viéramos a un alumno completar una tarea de evaluación, quizás no pudiéramos distinguir si dicha tarea era parte de una evaluación individual o nacional. No obstante, una vez que el alumno ha terminado su evaluación, lo que diferencia una evaluación individual de una nacional es la forma en que se utilizan los resultados.

En una evaluación nacional, los datos sobre el desempeño de los alumnos se agregan, generalmente para todos los alumnos del país o de un grupo representativo del mismo, y quizás incluso de varios grupos (por ejemplo, niñas, niños, alumnos de las zonas urbanas y rurales). Cuando se suman los resultados de esta forma, se considera a los desempeños como resultados del sistema o, utilizando otro término de origen económico, como indicadores.

Para que se les considere indicadores, los datos deben reunir ciertas características, a saber:

- Un indicador es cuantificable, es decir, que representa algún aspecto del sistema educativo de una forma numérica.
- Cada valor de un indicador se refiere sólo a un momento o período de tiempo.
- Una cifra estadística se considera indicador sólo cuando hay un estándar o criterio contra el que se puede contrastar. El estándar puede consistir en una comparación entre diferentes jurisdicciones con referencia a una norma (sincrónica); una comparación con referencia a sí misma (diacrónica) con valores

de indicadores del mismo sistema educativo obtenidos en diferentes momentos en el tiempo; o una comparación con referencia a un criterio con un ideal u objetivo preestablecido.

- Un indicador proporciona información sobre aspectos de un sistema educativo que los políticos, los reformadores o el público en general consideran importantes. A veces puede ser fácil obtener consenso entre las diferentes partes interesadas sobre lo que es importante; otras veces, puede no serlo.
- Un indicador es realista en el sentido de que se basa en información obtenida teniendo en cuenta limitaciones financieras y de otro tipo.
- Un indicador describe condiciones que se pueden mejorar..
- La información para los indicadores se recoge con la frecuencia suficiente para poder hacer un seguimiento del cambio.
- Un indicador permite examinar las distribuciones entre las subpoblaciones del caso (por ejemplo, de acuerdo a la edad, el sexo o el grupo social) (Greaney y Kellaghan, 1996).

La selección de indicadores para describir el estado del sistema educativo debería basarse en un modelo, bien sea explícito o implícito, de cómo funciona el sistema educativo (Burnstein, Oakes y Guiton, 1992). Aún más, el conjunto de indicadores incorporados en el modelo debe reflejar la naturaleza multifacética de la educación en toda su complejidad (Bottani y Tuijnman, 1994) al tiempo que ser lo suficientemente amplio para describir las dimensiones importantes del sistema. El modelo, por su parte, debe facilitar el contexto para interpretar el significado de los indicadores, describir cómo se relacionan con otros aspectos del sistema educativo (y quizás con otros sistemas sociales y económicos) y sugerir cómo es probable que respondan a distintos tipos de manipulaciones.

El modelo de sistema educativo en que se basan los sistemas de indicadores de la mayoría de los países en la actualidad, es uno en que domina la investigación sobre la eficacia de la escuela y se basa en la consideración de insumos, procesos y resultados. Los insumos son los recursos disponibles en el sistema—por ejemplo, edificios, libros, número y calidad del profesorado, y características educativas relevantes de los alumnos tales como las condiciones socioeconómicas de sus familias, comunidades y regiones. Los procesos son las formas en que las escuelas utilizan sus recursos, por ejemplo, en actividades de currículo y de instrucción. Los resultados son todo lo que la escuela trata de conseguir y comprenden los logros congnoscitivos de los alumnos y también los que no son cognitivos, como los sentimientos y las actitudes positivas y negativas que desarrollan los alumnos respecto a sus actividades, intereses y valores.

La elección de un modelo es importante ya que éste trata de identificar, en las escuelas, los factores que ejercen una influencia sobre el aprendizaje del alumno. Es por ello que debemos preguntarnos qué pasaría si los factores identificados como insumos y procesos en los modelos de insumos-resultados de tendencia funcionalista no son los únicos que ejercen una influencia, o no están

adecuadamente especificados. Modelos alternativos nos hacen pensar que esto puede ser precisamente lo que pase. Estos modelos, por ejemplo, señalan factores en el sistema educativo que no están adecuadamente representados en modelos de insumos-resultados y esto no es importante sólo por sí mismo sino que hay evidencia empírica de que influyen en el rendimiento escolar. Entre tales factores se encuentran los siguientes: el aspecto comunitario de las escuelas que se centra en la escuela como una "pequeña sociedad", una organización en la que las relaciones informales y duraderas se guían por un *ethos* común; las normas y prácticas de la escuela por medio de las cuales los alumnos entran en contacto con las asignaturas: el compromiso de los alumnos que comprende la participación, la conexión, el vínculo y la integración de estos últimos en el ambiente y la labor educativa de la escuela, y para los cuales las relaciones personales entre alumnos y maestros pueden funcionar como catalizador; y, por último, la participación de los padres manifestada en su contribución al aprendizaje escolar de los niños o en el establecimiento de una comunidad en torno a la escuela (Lee, Bryk y Smith, 1993).

Cualquier modelo de enseñanza que elijamos estará limitado en su capacidad para explicar el aprendizaje de los alumnos. Sin embargo, es importante estar consciente de la naturaleza de las limitaciones y darse cuenta de que éstas se reflejarán en los sistemas de indicadores ligados al modelo, al igual que en cualquier medida de mejora que se adopte con base en la información contenida en los indicadores.

USOS DE LOS DATOS DE UNA EVALUACIÓN NACIONAL

La recolección de datos para construir indicadores que reflejen los resultados del sistema educativo, así como la de otros datos que pueden ser necesarios para interpretar tales indicadores, es cara. Por esta razón, tiene sentido preguntar: ¿Para qué se van a utilizar los indicadores derivados de una evaluación nacional? Existen varias propuestas en la literatura sobre este asunto (Greaney y Kellaghan, 1996). A continuación se presentan ocho de estos usos, los cuales no agotan todos los objetivos de evaluaciones nacionales que se han perseguido en algunos países.

Informar la política

Este es un uso muy general e incluye otros. Se sostiene que el tipo de información disponible a través de las evaluaciones nacionales proporciona una base objetiva y, por consiguiente, más sólida para la toma de decisiones políticas que otros factores que contribuyen a la toma de decisiones tales como las tendencias personales de los ministros de educación o altos funcionarios, los intereses creados de los administradores escolares o los sindicatos de maestros, o los datos anecdóticos presentados por las empresas, los periodistas o los políticos. Sea como fuese, sería ingenuo pensar que las políticas se formulan al margen de tales factores, incluso cuando hay información disponible de la evaluación nacional.

Sabemos relativamente poco sobre cómo los datos obtenidos por medio de las evaluaciones nacionales influyen sobre las políticas de los países en los que tienen

lugar tales evaluaciones, pero es posible suponer que suceda algo similar al uso de la evidencia proveniente de la investigación[2]. En algunos casos, se puede actuar directamente sobre la información, por ejemplo, cuando se toma una decisión de política para aumentar el tiempo dedicado a un área del currículo o para introducir un nuevo tema en el programa nacional de estudios. Se han relatado ejemplos de tales decisiones como respuestas a los resultados de evaluaciones internacionales (Kellaghan, 1996a).

No obstante, es probable que la información procedente de una evaluación nacional (como sucede con la investigación) influya con más frecuencia en la política, introduciéndose en el campo de la toma de decisiones como parte de una búsqueda interactiva de conocimiento. En este caso, los datos de los indicadores son sólo una fuente de información y experiencia en el complejo proceso de la formación de políticas. Por otro lado, la información derivada de una evaluación nacional puede no estar realcionada directamente con las decisiones pero puede servir para "iluminar" el proceso de elaboración de políticas, proporcionando ideas sobre el alcance y las causas de los problemas y sobre soluciones apropiadas para resolverlos.

Dar seguimiento a los estándares

El seguimiento de los estándares a lo largo del tiempo suele mencionarse como una función importante de la información procedente de una evaluación nacional. A raíz de esto, muchos países recogen información de una forma cíclica—con una periodicidad que va desde uno hasta diez años. Si bien es cierto que la idea de dar seguimiento a los estándares es sumamente atractiva, sobre todo cuando tantos críticos denuncian—generalmente con base en datos inadecuados—el descenso en el rendimiento escolar, la información procedente de una evaluación nacional se encuentra limitada, desgraciadamente, en lo que puede comunicarnos sobre las tendencias del rendimiento escolar que podría servir para responder a los críticos.

En primer lugar, el tipo de rendimiento medido por las evaluaciones nacionales puede no ser el que interesa al empresario que se queja del sistema escolar. Cada vez más, los críticos hablan del desajuste entre los conocimientos y las habilidades desarrolladas en las escuelas y el tipo de destrezas cognitivas y no cognitivas necesarias para lograr el éxito social y económico en el mundo contemporáneo: aptitudes de razonamiento de alto nivel, análisis, resolución de problemas, pensamiento crítico, adaptabilidad, trabajo en equipo y flexibilidad. Probablemente, las críticas continuarán hasta que se demuestre que las evaluaciones nacionales pueden proporcionar evidencia de tales resultados.

Un segundo problema que surge al querer emplear la información de las evaluaciones nacionales para medir tendencias a lo largo del tiempo es que ésto es extremadamente difícil, incluso se puede decir que es imposible de llevar a cabo por razones técnicas, entre otras (Goldstein, 1996). Desde luego, los cambios en el programa de estudios, en los instrumentos de medida y en la composición de la población escolar hacen muy difícil comparar el desempeño escolar evaluado en

diferentes momentos.

Teniendo en cuenta estas dificultades, no debe extrañarnos encontrar diferencias en la interpretación de los datos de una evaluación nacional. Por ejemplo, al hablar de la evaluación nacional realizada en Chile, Olivares (1996) afirma, tras comparar los resultados de 1988 y 1992, que ha habido una mejora en español y en matemáticas. Himmel (1996), por su parte, dice que, debido a que no se tomó ninguna medida para garantizar que las pruebas se pudieran comparar de un año para otro, no se pueden contrastar los cambios en el desempeño escolar a lo largo del tiempo.

Decidir sobre la asignación de recursos

Algunas evaluaciones nacionales han sido concebidas para ayudar a los administradores de la educación a tomar decisiones sobre la asignación de los recursos. Así, los resultados de la evaluación nacional de Chile fueron utilizados para identificar las escuelas que tenían bajos resultados y a las cuales se les facilitaron entonces recursos. De las 8.000 escuelas de educación básica del país, se identificaron unas 900 de las zonas rurales y de las urbanas más pobres para la citada intervención (Olivares, 1996).

Aunque la utilización de los datos de una evaluación nacional para distribuir recursos es, en principio, atractiva, trae consigo al menos dos desventajas. Por un lado, exige una evaluación de todas las escuelas, no solamente de una muestra. Y, por otro, existe el peligro de que los maestros manipulen los datos obtenidos en la evaluación (por ejemplo, al comunicar el número de niños de familias menos favorecidas) para ampliar sus posibilidades de obtener recursos adicionales.

Introducir estándares realistas

Los resultados de una evaluación nacional, y aún más los de una internacional, pueden crear bastante conmoción en la comunidad educativa de un país. Esto es lo que sucedió recientemente en Sudáfrica cuando se dieron a conocer los resultados del desempeño escolar en matemáticas y ciencias en el Tercer Examen Internacional de Matemáticas y Ciencias (Third International Mathematics and Science Survey, TIMSS), organizado por la Asociación Internacional para la Evaluación del Rendimiento Educativo (International Association for the Evaluation of Educational Achievement, IEA). Se consideró que los desempeños de los alumnos blancos y negros eran muy pobres de acuerdo a estándares internacionales.

Existe el peligro de que la información sobre los resultados obtenidos por los alumnos (cuando se comparan con los de otros países, pero no sólo en ese caso) sea desalentadora para países menos desarrollados económicamente y en los cuales los recursos disponibles para la educación se encuentran mucho más limitados que en países más ricos y con mejores resultados académicos. Si el logro de información sobre los resultados de los estudiantes lleva al fatalismo o al pesimismo, no habrá cumplido con su cometido. Debemos sopesar este riesgo contra el valor de obtener

una base realista para hacer frente a los problemas del sistema educativo.

La participación en una evaluación internacional tiene la ventaja de que expone a los participantes a una gran variedad de destrezas a las que probablemente se tendrá acceso en un sólo país. Puede así contribuir de un modo significativo a la consolidación de la capacidad que pueda más adelante ser utilizada para diseñar y poner en práctica evaluaciones nacionales. Antes de decidir sobre la participación en una evaluación nacional, sin embargo, se debería considerar una serie de puntos. Primero, a la luz de la relación entre el desarrollo socioeconómico de un país y el nivel de desarrollo de su sistema educativo, ¿proporcionarán los demás países una base razonable de comparación?. Segundo, ¿existen otras demandas en el sistema educativo (por ejemplo, el establecimiento de escuelas primarias en áreas donde éstas son insuficientes) tan obvias que sería muy difícil de justificar el gasto en una evaluación internacional? Y, tercero, si se decide tomar parte en una evaluación internacional, ¿está claro de qué recursos se dispone para hacer frente a las insuficiencias del sistema educativo que se pongan de manifiesto en la evaluación? Si no se puede tomar ninguna medida para resolver esas insuficiencias, no parece que tenga sentido gastar considerables sumas de dinero en obtener la información necesaria.

Identificar aspectos relacionados con el rendimiento

El valor de las estadísticas descriptivas en sí mismas es sumamente limitado. No obstante, su valor puede verse ampliado considerablemente si se ponen en contexto. De hecho, cierta contextualización debería ser un requisito mínimo a la hora de realizar una interpretación. Teniendo esto presente, muchas evaluaciones nacionales recogen información que no sólo pone los resultados obtenidos por los alumnos en contexto, sino que incluso indica posibles explicaciones de la variación de los resultados de rendimiento.

Existen muchos casos de evaluaciones nacionales en los que se identificaron aspectos relacionados con el rendimiento. Por ejemplo, en una evaluación nacional en Colombia, se descubrió que una serie de factores estaban relacionados con el rendimiento escolar; entre ellos, el énfasis que los maestros ponían en ciertas áreas del currículo, la formación de los propios maestros, las condiciones de vida de los alumnos, la proporción de niños y libros de texto y el sexo de los alumnos (las niñas tenían mejores resultados en lengua española; los niños se destacaban más en matemáticas) (Rojas, 1996).

Dos cuestiones son importantes en el contexto de la identificación de los aspectos relacionados con el rendimiento. En primer lugar, existe el peligro de que se intente recolectar demasiada información, cruzando la frontera de lo que tradicionalmente se considera investigación. En segundo lugar, se corre el riesgo de que se interpreten las correlaciones como si se tratara de relaciones de causa y efecto, no contando con suficiente evidencia que las sustenten. Aún más, interpretaciones en esta línea de causa y efecto pueden llevar a la toma de decisiones sobre intervenciones. No

obstante, el que haya correlaciones entre las variables contextuales y el rendimiento escolar no proporciona certeza absoluta sobre la existencia de relaciones causales, ni de los posibles efectos de la manipulación de las variables. Esto no quiere decir que, con base en otros datos relevantes y juntamente con información sobre el indicador, no se pueda justificar la propuesta de algún tipo de acción. De cualquier manera, tales acciones deben realizarse con precaución y preferiblemente con carácter piloto.

Dirigir los esfuerzos de los maestros y elevar los rendimientos de los alumnos

Suponiendo que las escuelas considerarían que el contenido de la evaluación nacional reflejaba resultados importantes del aprendizaje, Tailandia introdujo un sistema de evaluación que comprendía resultados que no se solían detectar en las escuelas del país (Greaney y Rojas, 1996). Las autoridades tailandesas tenían probablemente razón al esperar que el contenido del currículo de los exámenes nacionales, así como la información sobre el desempeño de los alumnos que tales exámenes proporciona, acabaría por mudar el comportamiento de los maestros. Además, si los maestros comienzan a recalcar y a dedicar tiempo a nuevas áreas del currículo, es de esperar que también se opere un cambio en el rendimiento escolar.

Esta forma de pensar se basa en el hecho ampliamente admitido de que, si se considera que una evaluación es importante, es probable que la enseñanza y el aprendizaje acaben por alinearse con lo que se evalúa. Una situación en la que se considerará que tal información es importante es cuando se trata de una evaluación con importantes consecuencias, es decir, cuando al desempeño se vinculan ciertas sanciones, bien sea a los maestros o a los alumnos.

En la evaluación del currículo nacional en Inglaterra y Gales, se somete a prueba a todos los alumnos de determinados cursos en todas las escuelas, en parte, con el fin de guiar los esfuerzos de los maestros y en parte con vistas a mejorar la responsabilidad de las escuelas. Por su parte, la evaluación nacional en los Estados Unidos fue diseñada como una discreta medida del sistema educativo, que se limita a describir lo que saben y pueden hacer los alumnos, sin tratar de ejercer ninguna influencia directa sobre lo que sucede en las escuelas. Este enfoque, sin embargo, parece estar cambiando.

Al diseñar una evaluación nacional, se deberían sopesar seriamente las consecuencias de escoger uno de los dos modelos, el que acarrea importantes consecuencias al vincular sanciones y desempeño, y el otro que no. Por un lado, asociar decisiones importantes al rendimiento aumenta la posibilidad de que la evaluación afecte a la práctica escolar. Por otro lado, esta opción puede originar una serie de problemas. Primero, se considerará que el área que se evalúa constituye un indicador importante de lo que tiene valor en la educación, lo cual puede llevar a un cierto descuido en la enseñanza de otras áreas importantes del currículo. Segundo, con el fin de mejorar su desempeño en la evaluación, las escuelas pueden

adoptar estrategias tales como negar la admisión a los alumnos de bajo rendimiento, animar a tales alumnos a abandonar la escuela, impedirles presentarse a examen, retenerles en ciertos cursos o fomentar su ausencia el día del examen. Tercero, las escuelas se esforzarán bastante por prepararse para los exámenes. Esto puede incluir capacitarse para la prueba, elegir objetivos basados en preguntas de las pruebas y enseñar de acuerdo con los mismos y familiarizar a los alumnos con preguntas similares a las contenidas en el examen. A medida que se acentúa el paralelo entre los procesos de instrucción y las preguntas del examen, el desempeño del alumno en la prueba mejorará, efectivamente, pero no será posible decir que la mejora en la calificación de la prueba sea reflejo de un aumento de los conocimientos y destrezas. En efecto, la evidencia empírica indica que una mejora en el desempeño en medidas en las que los alumnos son capacitados intensamente no va unida a una mejora en el desempeño en otras medidas de rendimiento (Madaus y Kellaghan, 1992).

Promover responsabilidad

Adjudicar consecuencias a las pruebas, tal como dijimos en el apartado anterior, crea problemas de responsabilidad. El uso de datos de una evaluación nacional con fines de rendición de cuentas tiene además otras complejas implicaciones. A quién se considere responsable dependerá en cierta medida del nivel al que se analice el desempeño. Así, en el sistema de evaluación británico, en que se informa sobre el desempeño al nivel individual de cada escuela, la responsabilidad recae sobre las escuelas. Sin embargo, cuando no se pueden identificar maestros o escuelas individuales, o cuando sólo una muestra de la población, y no su totalidad, participa en la evaluación, no se puede determinar qué maestros o escuelas muestran bajo rendimiento y, por tanto, no se les puede exigir responsabilidad por ello.

Cuando sólo se presenta la información al nivel del sistema educativo, parecería lógico pedir responsabilidad al gobierno o al ministerio. Es por esta razón que los ministerios pueden estar poco dispuestos a participar en una evaluación nacional o a hacer públicos sus resultados una vez realizada la misma.

Olivares (1996) ha trazado las líneas generales de algunos efectos negativos de una evaluación nacional, el Sistema Nacional de Medición de la Calidad de la Educación (SIMCE), que están unidos al tema de la responsabilidad en Chile:

> Hay maestros que desarrollan sus programas de enseñanza como si se estuvieran "preparando para el SIMCE" e incluso hay casos en los que han tratado de "ayudar" a sus alumnos de formas no ortodoxas. Aún más grave es el hecho de que haya centros en los que se ha pedido a los maestros que se marchen cuando los resultados obtenidos no se ajustan a las expectativas optimistas del centro. Del mismo modo, hay directores de centros escolares que no comunican a los gobernadores los resultados bajos por si éstos pudieran crear dificultades. Finalmente, existen diferentes secciones de las

comunidades sociales y escolares que tienden a crear una "liga de escuelas buenas y malas (pág. 133).

Aumentar la concientización pública

Si bien en ocasiones puede ser oportuno a corto plazo que el Ministerio de Educación fije límites a la cantidad de información que va a poner a disposición del público, a largo plazo las ventajas de un sistema abierto de información probablemente sobrepasarán las desventajas a corto plazo. Por una razón: la ampliación del conocimiento del público sobre los asuntos educativos es probable que aumente su apoyo a la reforma educativa y a los fondos que tal reforma necesita.

Aunque existen estudios que apoyan los anteriores usos de la evaluación , es difícil saber, en la práctica, cuáles de ellos ocupan los pensamientos de los responsables de las políticas o cúales son los que les animan a embarcarse en una evaluación. No cabe duda de que algunas de las razones aducidas por Nwana (1996) al estudiar la participación de los países africanos en evaluaciones nacionales, no están claras. El deseo de Uganda de realizar un seguimiento a lo largo del tiempo y el de Burkina Faso de identificar las áreas conflictivas del currículo están suficientemente claros. Sin embargo, el empeño de Cabo Verde por evaluar la eficiencia del sistema educativo, el de Lesoto por medir la "calidad" de la instrucción, el de Madagascar por mejorar la "calidad" de la educación y el de Mali y Mozambique por juzgar los efectos de las innovaciones para incrementar la "calidad", habrían de detallarse más antes de que se pudiera empezar a planificar.

DESCRIPCIÓN DEL ESCENARIO DE UNA EVALUACIÓN NACIONAL

Una vez que se ha decidido llevar a cabo una evaluación nacional, es importante hacer todo lo posible para asegurar que cuente con el apoyo suficiente, que funcione normalmente y que sus resultados se tengan en cuenta a la hora de tomar decisiones de política. Esto exige la participación desde el inicio de todos los interesados y, en la medida de lo posible, el logro de un consenso sobre la evaluación. Una forma de hacerlo es establecer un comité promotor o una junta independiente de gestión, en la que estén representados diversos intereses educativos y comunitarios, lo cual, si bien no es esencial para una evaluación nacional, algunos comentaristas lo recomiendan enérgicamente (Greaney y Rojas, 1996; Lapointe, 1990). Las funciones de este comité serían las de unir los más amplios fines sociales y políticos de la evaluación nacional con los aspectos más técnicos de su ejecución. El comité también daría categoría a la evaluación, contribuiría a asegurar que se prestara atención a las necesidades de los grupos influyentes en educación en el país y podría ayudar a suprimir los obstáculos administrativos y financieros que pueden comprometer o paralizar un proyecto de evaluación. El comité proporcionaría dirección general y promovería el conocimiento por parte del público y la discusión de los resultados, buscando el impacto de la evaluación en la política educativa.

Dado que las estructuras de poder político-educativo difieren entre países, los intereses representados en el Comité Nacional de Gobierno variarán de un país a otro. Obviamente, se debe dar representación a los encargados de la administración de la evaluación nacional, los responsables de su financiamiento, quienes tendrán en cuenta los resultados para desarrollar políticas, y aquéllos a quienes se encargará de ejecutar las reformas que puedan surgir de la evaluación, tales como los administradores de escuela y los maestros. En general, cuanto más homogéneo es el país y su sistema educativo, más fácil será establecer una evaluación nacional. Por ello, sería menos difícil establecer un sistema de evaluación en un país en el que existe una sola autoridad educativa y un sistema de educación uniforme que en uno que cuenta con muchas autoridades educativas (por ejemplo, en un sistema federal) y en el que no hay un currículo nacional (Nwana, 1996).

Entre los asuntos más importantes que el comité promotor debería abordar están los de identificar el propósito y fundamento de la evaluación nacional, decidir sobre el contenido y los niveles a los que se aplicará, preparar un presupuesto e identificar la forma de control presupuestario, seleccionar la institución o instituciones que llevarán a cabo la evaluación, determinar los términos de referencia y decidir sobre la forma de presentación y publicación de los resultados.

Greaney y Rojas (1996) indican que las evaluaciones nacionales en Chile, Colombia y Tailandia se beneficiaron de un compromiso por parte de los ministerios de educación. En Colombia, la instauración de un sistema de evaluación estuvo probablemente facilitada, además, por la promulgación de una ley en 1986 que establecía una División, dentro del Ministerio de Educación, con la responsabilidad de evaluar la calidad de la educación en las escuelas (Rojas, 1996). No obstante, el compromiso del gobierno no se puede considerar una garantía de efectos positivos. En Inglaterra y Gales, el gobierno fue la fuerza promotora tras la evaluación del currículo nacional. Sin embargo, la evaluación provocó gran rechazo por parte de los maestros que estaban particularmente preocupados por la utilización de los resultados para pedir cuentas a escuelas y maestros sobre el desempeño de los alumnos. En un año, los maestros se negaron de hecho a cooperar en la administración de las pruebas, si bien las razones para ello fueron complejas. También surgieron problemas en la evaluación nacional en Egipto donde los gobiernos locales la vieron como una amenaza a la autonomía local y retrasaron su ejecución.

ELECCIÓN DE UNA INSTITUCIÓN PARA REALIZAR LA EVALUACIÓN NACIONAL

La ejecución de la evaluación nacional exige capacidad técnica, o acceso a tal capacidad, en una serie de actividades que incluyen la gestión del proyecto, el análisis del currículo, la elaboración de pruebas y cuestionarios, el muestreo, la impresión de los instrumentos, la recolección de datos, el procesamiento y análisis de los mismos y la preparación de informes con los resultados. Muchos Ministros

de Educación no buscan tal pericia más que entre su personal. Esto es lo que sucedió en Tailandia, donde se encargó de la evaluación nacional a la Oficina de Evaluación y Capacitación Educativa, una sección del Ministerio de Educación. En muchos otros países en vías de desarrollo, se pueden contratar para trabajar en el Ministerio a algunos de los más destacados educadores. Otras razones para basar la evaluación en el Ministerio son que el personal del mismo suele tener acceso a información actualizada para llevar a cabo el muestreo y que los inspectores escolares o miembros de las unidades de currículo o libros de texto tienen supuestamente una visión de los aspectos clave del sistema educativo. Si existe una estructura dentro del Ministerio para realizar tareas relacionadas con una evaluación nacional, ésta debe ayudar a asegurar que se preste adecuada atención a cada fase de la evaluación. Esta es la situación de Colombia, donde existen unidades separadas para ejecutar una evaluación (incluyendo muestreo, recolección y análisis de datos), para actividades de investigación (incluyendo diseño de instrumentos y análisis posterior) y difusión (es decir, comunicación de los resultados) (Rojas, 1996).

Existen también razones en contra de encargar al Ministerio de Educación la administración, por sí solo, de una evaluación nacional. Por un motivo: muchos ministerios carecen de la competencia técnica necesaria. Aún más, cuando un Ministerio realiza una evaluación nacional, tiende a demorarse en compartir los resultados con otros. Dado que es probable que los funcionarios del Ministerio tengan un interés creado en los resultados de una evaluación, puede que no manifiesten gran entusiasmo por centrarse en asuntos potencialmente difíciles o por hacer públicos los resultados desagradables. Por ejemplo, los resultados que ponen de manifiesto una prestación deficiente de un servicio educativo o un fracaso del sistema de educación formal para conseguir un fin particularmente delicado (tal como la igualdad del rendimiento de los diferentes grupos étnicos) pueden poner en un aprieto a los funcionarios del Ministerio y, aún más críticamente, a sus jefes políticos.

Se puede defender, por tanto, con argumentos sólidos la conveniencia de involucrar a una institución externa en la realización de una evaluación nacional. Esta postura se ve subrayada por el hecho de que es posible que los principales interesados en la educación consideren que la información facilitada por una organización no gubernamental o independiente de renombre es más objetiva y, por tanto, más admisible. A esto se añade que es más fácil encontrar competencia técnica dentro de los departamentos de una universidad o institutos de investigación independientes que dentro de los ministerios de educación, si bien los departamentos universitarios no suelen tener la capacidad para realizar trabajo de campo a la escala que exige una evaluación nacional. Fue por razones como las mencionadas que inicialmente se encargó la evaluación nacional en Chile a un organismo no gubernamental. En Inglaterra y Gales también, instituciones externas juegan un papel muy importante en la elaboración de los instrumentos mientras que las escuelas mismas son responsables de la administración local.

El empleo de una organización externa, sin embargo, no deja de tener problemas.

En anticipación de los mismos, se debe realizar un memorándum de acuerdo entre el comité promotor y la institución ejecutora antes de que comience el trabajo. El memorándum debe tratar temas como la financiación, los calendarios, las relaciones entre las dos partes y el uso permitido de los datos.

Una alternativa a confiar solo en un organismo interno o externo es encargar de la evaluación nacional a un equipo compuesto tanto de personal del Ministerio de Educación como de expertos técnicos y del currículo de fuera del Ministerio. Un arreglo de este tipo puede aprovechar las ventajas de ambos grupos y aumentar la posibilidad de aceptación general de los resultados de la evaluación. Son posibles varios consorcios. En Mauricio, el semiautónomo Sindicato Mauritano de Exámenes (Mauritian Examination Syndicate) acometió una tarea de desarrollo para la evaluación nacional en colaboración con otras instituciones nacionales, incluidos el Ministerio de Educación, Ciencia y Tecnología, el Instituto de Educación y su Centro de Desarrollo del Currículo. En Colombia, el Ministerio de Educación trabajó en cooperación con varias instituciones públicas y privadas. Para el diseño de los instrumentos se contrató al Servicio Nacional de Pruebas y, para la recolección y el análisis de los datos, al Instituto de Investigación SER (Rojas, 1996). En Chile, cuando se restablecieron las evaluaciones nacionales en 1988, el Ministerio de Educación contrató a una Universidad para que realizara la evaluación pero el personal del Ministerio trabajó con el de la Universidad durante un período de más de tres años, desarrollando aptitudes que podrían utilizarse más adelante. A partir de entonces, el personal de la Universidad prestó solamente servicios de consejería (Himmel, 1996).

Cuando no se cuenta con la competencia profesional necesaria dentro del país, puede ser indispensable contratar a expertos extranjeros. Existen numerosos ejemplos de ello. El análisis de datos para la evaluación nacional de Namibia fue dirigido por la Universidad de Harvard, mientras que la Universidad del Estado de la Florida ayudó con aspectos del muestreo. Cuando se hace uso de expertos extranjeros, éstos deben responder al comité promotor y ayudar en el desarrollo de la capacidad local para realizar futuras evaluaciones.

¿A QUIÉN EVALUAR?

Las deducciones sobre los resultados de un sistema educativo se basan en una evaluación de los rendimientos de los alumnos. Para realizar estas deducciones, sin embargo, no es necesario ni deseable involucrar en la evaluación a estudiantes de todas las edades y niveles. Es necesario tomar dos decisiones. Una se refiere al nivel de escolaridad que será evaluado y la otra, a la elección entre centrarse en alumnos de un curso específico o los de una edad en particular.

Elección del nivel de escolaridad

Los políticos desean obtener información sobre el conocimiento y las habilidades de los alumnos en momentos determinados de sus carreras educativas.

Prácticamente en todos los países que llevan a acabo evaluaciones nacionales, se somete a prueba a los alumnos de primaria. Por ejemplo, en Chile, se evalúa a los alumnos de cuarto porque es el último curso de un subciclo de la educación general básica (Himmel, 1996). En la mayoría de los países se realizan, además, evaluaciones nacionales en algún momento de la escuela secundaria, generalmente en los primeros cursos, cuando la educación todavía es obligatoria. Tomando de nuevo a Chile como ejemplo, los alumnos del nivel secundario son evaluados en el octavo curso, que es el último año de la educación obligatoria. La información sobre ambos niveles puede ser valiosa. Las evaluaciones de primaria pueden identificar deficiencias en un momento temprano del sistema educativo, poniendo de manifiesto la necesidad de emprender una acción correctiva. La información obtenida hacia finales de la escolaridad obligatoria, o en un momento en que gran parte de los jóvenes todavía asisten a la escuela, puede también ser útil si proporciona alguna indicación de qué tan bien preparados están los alumnos para la vida una vez abandonen la escuela. En muchos países en vías de desarrollo esto sucede a nivel de primaria.

Población definida por edad o curso

En algunas evaluaciones nacionales (por ejemplo, Chile y Escocia), sólo se tiene en cuenta el curso a la hora de definir la población para una evaluación nacional. Muchas evaluaciones nacionales e internacionales, sin embargo, utilizan en su definición tanto la edad del alumno como el curso. Por ejemplo, en el estudio de alfabetismo IEA, se definieron dos poblaciones: alumnos en el curso que contenía más niños de nueve años y alumnos en el curso que contaba con más niños de catorce años (Elley, 1992). Recientemente, en los Estados Unidos también se ha seleccionado el curso de la mayoría de los alumnos de una determinada edad (Johnson, 1992). Esta estrategia puede justificarse en los países industrializados, donde la norma es la promoción automática al final de cada curso, asegurando así un pronunciado vínculo entre el curso y la edad.

En los países en vías de desarrollo, especialmente en Latinoamérica y en la zona francófona del oeste africano, la relación entre la edad y el curso puede no ser tan estrecha debido a la gran variación de edades de entrada a la escuela y a normas que limitan la promoción. En esta situación, los alumnos de edad similar no se encontrarán en el mismo curso. Seleccionar una población con base en la edad en un sistema escolar de este tipo sería disruptivo ya que exigiría que los alumnos de varios cursos diferentes se sometieran a las pruebas al mismo tiempo. También sería difícil identificar el contenido de la prueba para la gama de rendimientos que sería de esperar demostraran tales alumnos. En vista de lo cual, existen razones sólidas para seleccionar por curso en vez de hacerlo por edad en las evaluaciones nacionales de los países en vías de desarrollo.

¿Población total o una muestra de la población?

Una vez que se llega a una decisión sobre la edad o el curso que proporcionarán

información para una evaluación nacional, el siguiente paso es ver si se evaluará a todos los alumnos de la edad o curso señalados o si sólo se elegirá una muestra para la evaluación. Encontramos ejemplos de ambos planteamientos. En Chile, Egipto, Inglaterra y Gales y Francia, se evalúa a todos los alumnos. En la mayoría de los países, sin embargo, más que a toda la población, se evalúa a una muestra de la misma. Ambos enfoques presentan ventajas y desventajas. La ventaja de usar toda la población es que se dispone de información sobre todos los alumnos y las escuelas, la cual se puede utilizar con fines formativos (como en Francia, donde los resultados se utilizan para diagnosticar los problemas de aprendizaje de los alumnos o en Chile, donde se utilizan para identificar las escuelas que necesitan recursos adicionales). Los resultados también pueden utilizarse con fines de identificar responsabilidades (como en Inglaterra y Gales). Sin embargo, a menos que se necesite información a nivel de alumno o escuela, el empleo de una muestra puede ser más apropiado. Existen, de hecho, varias ventajas ligadas a esta estrategia, incluyendo reducidos costos de recolección y análisis de los datos, mayor velocidad en el análisis y la comunicación de los mismos y mayor precisión debido a la posibilidad de supervisar más intensamente el trabajo de campo y la preparación de los datos (Ross, 1987).

Tanto si se decide evaluar a toda la población como si se opta por una muestra, es necesario identificar con precisión la población de interés. Ya que hay interés en el desempeño de los estudiantes de una determinada edad o curso en el país, puede parecer conveniente definir la población en términos de los alumnos. Sin embargo, es probable que esto no sea práctico. En la mayoría de los países, una institución central, como el Ministerio de Educación, no tendrá una lista con todos los alumnos que asisten a la escuela y que sería necesaria para seleccionar una muestra de probabilidad. E incluso si tal lista existiese, no sería eficiente o posible evaluar a una muestra de estos alumnos porque, si fuesen elegidos al azar, estarían dispersos en un gran número de escuelas, lo cual haría la recolección de datos difícil y cara. Dadas estas condiciones, se suelen escoger escuelas (es decir, grupos de alumnos) como población de la cual se van a tomar las muestras en la primera etapa.

Para elegir escuelas se necesita una lista actualizada de todas ellas junto con la información que sea relevante para la estratificación (por ejemplo, tamaño y localización). En algunos países quizás no se disponga de listas completas, o puede que éstas sean demasiado antiguas para ser útiles. Incluso cuando existe una lista de escuelas, se pueden tener buenas razones para no querer incluirlas todas en la población para una evaluación nacional. Se puede decidir, por ejemplo, excluir de la evaluación a las escuelas en las que se considera que los alumnos no son evaluables por dificultades del aprendizaje o por conocimiento limitado de la lengua en la que se realizará la evaluación. Se pueden agrupar escuelas muy pequeñas que no pueden proporcionar por sí solas un número suficiente de alumnos, para formar "pseudo escuelas", si bien esto puede ser complejo y caro desde el punto de vista administrativo. Debido al costo de la recolección de datos, se pueden excluir totalmente las escuelas situadas en zonas aisladas. Se deben reducir al máximo las exclusiones y proporcionar información sobre ellas en el informe de la evaluación

nacional.

Una vez que se ha elegido la población de todas las escuelas, el siguiente paso—a menos que se haya planeado evaluar a los alumnos en todas las escuelas—es seleccionar aquéllas en las que se van a evaluar los alumnos. Existe una variedad de estrategias para hacer esto y puede ser que la organización que realiza la evaluación nacional requiera de asistencia técnica para escoger la estrategia y decidir el número de escuelas y alumnos más apropiados. Se debe prestar gran atención a esta etapa ya que, si la muestra no representa adecuadamente la población, las afirmaciones sobre los niveles de rendimiento nacional de los alumnos no serán válidas.

Al realizar el muestreo de las escuelas, es frecuente estratificarlas de acuerdo con variables de interés, como la localización (zona del país, urbana o rural); el tipo (pública o privada); el grupo étnico de sus miembros y la confesión religiosa. Esto se debe a dos razones: la estratificación puede mejorar la confiabilidad de las estimaciones; también ayuda a asegurar que hay suficientes escuelas y alumnos en las diversas categorías, tales como urbana y rural, si se considera que las diferencias entre las escuelas (y los alumnos) de tales categorías son de interés. Para alcanzar números suficientes de escuelas y alumnos, puede ser necesario seleccionar un número mayor de las mismas dentro de ciertos estratos en vez de elegir justo el tamaño de muestra proporcional al número de escuelas (y alumnos) en el estrato. Cuando se seleccione por encima de lo necesario en un estrato, habrá de establecerse un sistema de ponderación al sumar las puntuaciones de los alumnos para estar seguros de que la contribución de los grupos a las estadísticas agregadas es proporcional a su tamaño en el total de la población, no a su tamaño en la muestra.

Después de la selección de las escuelas, la segunda etapa del muestreo exige que se tome una decisión sobre cómo se van a elegir los alumnos para la evaluación dentro de cada escuela. ¿Se evaluarán todos los alumnos de determinada edad? o, si existe más de una clase del nivel relevante, ¿se escogerá al azar solamente una clase o se seleccionarán varios alumnos de todas las clases? Si bien es cierto que la evaluación de clases intactas tiene obvias ventajas administrativas, la selección de alumnos de varias clases proporcionará una mejor estimación de los rendimientos de éstos en la escuela si se distribuyeron los alumnos entre las clases siguiendo diferentes criterios o si siguen diferentes currículos.

Determinar el tamaño óptimo de la muestra no es tarea sencilla. Lo que se pretende hacer es conseguir el nivel de precisión que se exige de los cálculos sin salirnos de los recursos disponibles para la recolección de los datos. Como guía general, se puede considerar el tamaño de la muestra utilizada en el TIMSS. En cada país se seleccionó una muestra de 150 escuelas y, dentro de cada una, se eligieron treinta alumnos. Sin embargo, según las circunstancias se puede necesitar una muestra mayor. Esto es probable que suceda si un país se encuentra dividido en una serie de regiones administrativas. También si se pretende poder comparar el desempeño en una serie de sectores del sistema educativo o si se esperan grandes diferencias en el rendimiento medio de unas escuelas a otras.

En una evaluación nacional no es necesario que todos los alumnos se sometan a la misma prueba. Se puede conseguir una amplia cobertura del currículo utilizando folletos con diferentes grupos de preguntas. Cada alumno responderá solamente a un folleto. Con este sistema, todavía se pueden calcular los promedios de las clases para cada tarea y agregarlos para obtener un resumen de las estadísticas. Este procedimiento es utilizado en los Estados Unidos, donde se administra, a muestras de alumnos, una séptima parte del número total de preguntas preparadas para cada curso. Esta manera de realizar el muestreo, llamada "muestreo de matrices" (matrix sampling), permite cubrir secciones del currículo mucho más amplias y puede resultar más rápida que administrar la misma prueba a todos los alumnos. No obstante, los requisitos técnicos y logísticos, incluyendo la impresión de muchas formas diferentes de una misma prueba, empaquetarlas y administrarlas y combinar los resultados de las pruebas, puede ser intimidante, especialmente en la primera evaluación nacional de un país.

La cooperación de la escuelas (las cuales pueden decidir, en la mayoría de los países, si desean o no participar en una evaluación nacional) es necesaria si se quiere lograr la muestra diseñada, lo cual, a su vez, es imprescindible si se van a hacer generalizaciones sobre el sistema educativo en general. Inevitablemente, uno se verá enfrentado con el tema de cómo tratar la falta de respuesta, debida bien sea a la negativa a cooperar de una escuela o a la ausencia del alumno el día del examen. No existe ninguna cifra absoluta de participación que se pueda considerar segura de que la muestra de las escuelas y los alumnos que responden representa adecuadamente el desempeño de la población meta en su conjunto. Cualquier grado de falta de respuesta puede ser motivo de preocupación. No obstante, si somos realistas, debemos aceptar que por diversas razones no todas las escuelas invitadas a participar lo harán. Como guía general sobre lo que se puede considerar una tasa aceptable de falta de respuesta, se pueden usar los estándares fijados para el TIMSS. En ese estudio internacional, la tasa mínima admisible de respuesta por parte de la escuela, antes de la utilización de escuelas suplentes, se fijó en el 85 por ciento. La misma cifra se utiliza para la tasa de respuesta de los alumnos dentro de las escuelas (Martin y Kelly, 1996).

Se pueden utilizar escuelas suplentes para asegurar un tamaño adecuado de muestra. Sin embargo, ellas no incrementan la tasa de respuesta y no pueden tomarse como prueba de que se ha suprimido cualquier posible tendencia derivada de la decisión de las escuelas de no participar. El informe de una evaluación nacional deberá contener información sobre las tasas de respuesta de las escuelas y los alumnos y la medida en la que se usaron escuelas suplentes.

Cuando se utiliza un diseño de muestreo complejo—que acarrea procedimientos tales como estratificación, agrupación y ponderación—para seleccionar a los participantes en un estudio, se deben tener en cuenta en los análisis los efectos del diseño en el error de muestreo. De lo contrario, el cálculo de la verdadera variabilidad del muestreo estará sesgada. Reducir este sesgo no es sencillo. Una

forma de calcular los errores de muestreo con datos de un examen a gran escala cuando no se dispone de las fórmulas, es usar un procedimiento de "navaja". Este procedimiento requiere que se hagan los cálculos de estadísticas sobre los datos de las muestras en su totalidad; a continuación, se dividen los datos en grupos y se realizan operaciones con grupos reducidos de datos en los cuales se omiten, uno tras otro, los subgrupos. Entonces se calcula la media de las submuestras y se toma su varianza como una aproximación de la estimación de las muestras. También existe un método alternativo para calcular el error implicado en un modelo de multinivel. Se considera estadísticamente más eficiente y tiene la ventaja de que proporciona información sobre la variación entre escuelas (Woodhouse y Goldstein, 1996).

¿QUÉ SE EVALÚA?

En la identificación del conocimiento y las habilidades que se van a examinar en una evaluación nacional entran en juego consideraciones políticas y técnicas. El papel de los factores políticos es evidente en la necesidad de seleccionar un contenido que responda a las necesidades de información de los políticos clave. Las consideraciones técnicas se manifiestan al decidir lo que es técnicamente posible medir y al evaluar el costo y los requisitos logísticos. En la práctica, es probable que las tensiones que hay que afrontar surjan entre las necesidades de información, los objetivos y las ambiciones de aquéllos que encargan la evaluación nacional y la capacidad de la evaluación para acomodarlos, teniendo en cuenta las limitaciones financieras, administrativas y de tiempo.

La información que se suele acumular en evaluaciones nacionales se puede dividir en tres categorías principales. Primero, todas las evaluaciones miden resultados cognoscitivos de instrucción—específicamente, la competencia en ciertas áreas del currículo. Segundo, admitiendo el enfoque de que la educación no debe limitarse al desarrollo de habilidades cognoscitivas, muchas evaluaciones nacionales e internacionales reúnen datos sobre resultados no cognoscitivos, los cuales incluyen el concepto de uno mismo, actitudes y valores. Tercero, la mayor parte de las evaluaciones nacionales también recogen información contextual sobre variables de antecedentes tales como factores escolares y no escolares que pueden influir en el rendimiento del alumno.

Resultados cognoscitivos

Todos los países que realizan evaluaciones nacionales examinan la lengua materna de los alumnos y las matemáticas. A veces se incluyen las Ciencias y, en un número menor de países, un segundo idioma, arte, música y estudios sociales (Kellaghan y Grisay, 1995). La atención prestada al lenguaje y a las matemáticas es una indicación de la importancia de estas asignaturas para la educación básica y merece una consideración seria por parte de cualquier país que desee embarcarse en una evaluación nacional, ya que la principal inquietud de tal evaluación debería ser la de recoger datos que proporcionen información sobre la medida en la que se están logrando importantes objetivos del currículo.

Una vez que se ha escogido un área específica del currículo para la evaluación, es necesario asegurarse de que las técnicas utilizadas para obtener la información sobre el rendimiento escolar cubran el currículo con gran amplitud. Esto es para que los resultados de una evaluación nacional proporcionen un retrato preciso del desempeño de los alumnos en áreas importantes del pretendido currículo, identificando los puntos fuertes y débiles del mismo logrado tal como se refleja en los perfiles de los resultados de los alumnos.

No sólo se debería abarcar todo el currículo, sino que se deberían seleccionar muestras de diversos campos con detalle suficiente para permitir que se realicen deducciones sobre la medida en la que se está enseñando y aprendiendo cada área en las escuelas. Así, el Tercer Examen Internacional de Matemáticas y Ciencias (TIMSS) evaluó seis áreas del currículo de matemáticas en las escuelas de primaria: números enteros; fracciones y proporcionalidad; medición, cálculo y sentido numérico; presentación de los datos, análisis y probabilidad; geometría; y patrones, relaciones y funciones (Mullis, Martin, Beaton, González, Kelly y Smith, 1997) y cuatro áreas de ciencias: ciencia de la tierra, ciencias de la vida, ciencias físicas y temas del medio ambiente y la naturaleza de la ciencia. La información diferencial que se puede obtener asegurando que se realiza un muestreo adecuado para cada área de contenidos, proporciona una buena base para determinar qué áreas del currículo se están ejecutando y logrando adecuadamente, y cuáles no.

Resultados no cognoscitivos

El desarrollo de los alumnos tiene muchas facetas y las escuelas, además de promover el desarrollo en las áreas cognoscitivas, perseguirán también aumentar el desarrollo de las habilidades personales y sociales de los alumnos, como son las relativas al concepto de uno mismo, la cooperación, el liderazgo, la innovación, la confianza en sí mismo, la motivación y la independencia.

Teniendo presentes consideraciones como éstas, algunas evaluaciones nacionales han incorporado mediciones de resultados no cognoscitivos. La evaluación nacional colombiana incluyó mediciones de las actitudes de los alumnos respecto a la escuela, asignaturas y maestros, creatividad, autoestima y valores democráticos, además de medidas de español y matemáticas (Rojas, 1996). En Chile, se administraron cuestionarios para evaluar la opinión que los alumnos tenían de sí mismos, sus actitudes respecto a la escuela y el aprendizaje, las relaciones con sus compañeros y sus relaciones sociales, su orientación profesional y la adquisición de valores (Himmel, 1996). A nivel internacional, el estudio de alfabetización de la IEA evaluó las actitudes de los alumnos respecto a la lectura examinando la medida en que se leían voluntariamente libros, tiras cómicas y periódicos, sus preferencias respecto a libros de lectura y la cantidad de estímulo que los alumnos recibían para leer y utilizar la biblioteca (Elley, 1994). Si bien los resultados no cognoscitivos resultarían ser un aspecto importante de la educación, se ha demostrado que su medida e interpretación son problemáticas y, en el caso de Chile, se consideró que

su uso no tuvo éxito (Himmel, 1996).

Variables de antecedentes

Ya hemos hecho referencia a la necesidad de tener un modelo de sistema educativo en mente cuando se realiza una evaluación nacional. Una parte importante de tal modelo será, probablemente, la información sobre las variables de antecedentes, tanto en la escuela como fuera de ésta, las cuales se utilizarán en la interpretación de los datos sobre el desempeño de los alumnos en las variables cognoscitivas y en las que no lo son.

En un intento por obtener información que ayudara a entender los factores que determinan el logro de un mínimo de objetivos de aprendizaje, el Programa de Evaluación del Rendimiento Escolar (PER) que se llevó a cabo en Chile entre 1981 y 1984 obtuvo datos sobre cinco grupos de variables de antecedentes. Estos grupos eran los siguientes: (i) antecedentes de los alumnos; (ii) maestros (formación, experiencia) y variables del aula; (iii) directores de escuela y escuelas (que incluye las expectativas de aprendizaje, administración y disciplina); (iv) ambiente local (que incluye el nivel socioeconómico, los recursos educativos); y (v) instituciones (estructura del sistema educativo, políticas educativas, financiamiento) (Himmel, 1996).

También en Tailandia se intentó identificar los factores relacionados con el rendimiento escolar de los alumnos, tales como el nivel socioeconómico, el tamaño de la escuela, la repetición de curso y el acceso al preescolar (Chinapah, 1992). En Colombia, se recogieron datos sobre el sexo de los alumnos, su historial educativo (incluyendo la repetición de curso), la duración de la jornada escolar, el acceso a libros de texto, el tiempo dedicado a los deberes y a ver la televisión, las características del maestro—tales como el nivel de educación formal y cuánta capacitación en servicio habían recibido—y factores de antecedentes del hogar (por ejemplo, instalaciones en la casa, tamaño de la familia, expectativas educativas y de trabajo, calidad de la vivienda) (Rojas, 1996).

INSTRUMENTOS DE EVALUACIÓN

Variables de rendimiento

Se puede señalar una serie de aspectos generales sobre la elaboración de los instrumentos de evaluación, tarea bastante técnica que probablemente se encargue a la organización responsable de llevar a cabo la evaluación. Primero, el contenido de los instrumentos debe ser coherente con los objetivos generales de la evaluación. Si uno de los objetivos, por ejemplo, es medir la competencia en los ámbitos matemáticos del cálculo, los conceptos y las aptitudes de resolución de problemas, el diseño del instrumento de evaluación debe asegurar que cada uno de los tres campos se evalúan adecuadamente. Segundo, siempre se deben tener presentes las

consideraciones de política. Esto supone que la cobertura del currículo en los instrumentos debe ser tal que permita realizar deducciones sobre la medida en la que se están logrando los objetivos del currículo en las escuelas.

Los documentos que especifican el currículo o programa de estudios, cuando se dispone de ellos, serán una fuente importante para el desarrollo de los instrumentos. Estos documentos, sin embargo, puede que no proporcionen detalle suficiente o indicación de la importancia relativa del contenido del currículo, en cuyo caso se tendrá que recurrir a otras fuentes, como los libros de texto y la experiencia del maestro.

La experiencia nos dice que un cuadro de especificaciones puede facilitar en gran medida el desarrollo de instrumentos de evaluación de selección múltiple o de respuestas breves (Bloom, Madaus y Hastings, 1981). Un cuadro típico consiste en un eje horizontal que enumera las áreas del contenido que se van a evaluar (por ejemplo, aspectos del currículo de matemáticas, tales como números enteros, fracciones, medición para un curso determinado) y un eje vertical que presenta de una forma jerárquica las aptitudes intelectuales o el comportamiento esperado en los alumnos (por ejemplo, cálculo, entendimiento). Se forman celdas en la intersección de los dos ejes. A quienes elaboran las pruebas, les corresponde asignar puntos o preguntas de la prueba a cada una de las celdas según sus percepciones de la importancia relativa del objetivo que cada celda representa. Se dejan las celdas vacías si se considera que el objetivo es inadecuado para un área del contenido en particular. El Cuadro 1 muestra un ejemplo de cuadro de especificaciones creado para un currículo de matemáticas para los cursos medios de la escuela primaria en Irlanda.

Si se pretende repetir una evaluación nacional más adelante para realizar un seguimiento de las tendencias del rendimiento a lo largo del tiempo, se tendrá que utilizar de nuevo la misma prueba (o una parte de la misma). En este caso, el instrumento de evaluación no debería hacerse público y deberían recogerse todas las copias del mismo inmediatamente después de la realización del examen. Los ejemplos de preguntas utilizadas en el procedimiento de la evaluación sí se pueden hacer públicos de forma que el personal de las escuelas sepa lo que se espera de los alumnos.

Tipo de prueba

La mayor parte de las evaluaciones nacionales e internacionales confían en gran medida, para sus instrumentos, en el formato de examen de selección múltiple. Un ítem de selección múltiple suele consistir en una afirmación, instrucción, o pregunta seguidas de una serie de respuestas alternativas, una de las cuales es la correcta. Entre las ventajas de este tipo de preguntas se encuentran la rapidez de la respuesta, la facilidad de responder o corregir, la objetividad al calificar, el potencial de cubrir una parte considerable de un área de contenidos y la gran seguridad o coherencia (Frith y Macintosch, 1984). Si se dispone de aparatos de lectura óptica,

las hojas de respuestas de selección múltiple se pueden calificar y procesar rápidamente, si bien en países donde la mano de obra es barata, la calificación manual puede ser todavía una alternativa. Cuando se utilizan computadoras en el tratamiento de los datos de selección múltiple, es fácil proporcionar información detallada sobre características de preguntas concretas, de objetivos y niveles de rendimiento de los alumnos—clasificados, por ejemplo, por curso, edad, sexo, grupo étnico o lingüístico, si se ha conseguido información sobre estas variables en el examen.

Cuadro 1. Cuadro de especificaciones para la prueba de matemáticas (cursos medios de primaria)

Los exámenes de selección múltiple también tienen aspectos negativos. Su elaboración lleva mucho tiempo y es cara. No se pueden utilizar para evaluar aspectos importantes del currículo, tales como la fluidez oral, la escritura y las destrezas prácticas. También han sido objeto de críticas por enfatizar demasiado los hechos a costa de determinar la comprensión que tiene el alumno del contenido que se evalúa.

Las numerosas desventajas de los exámenes de selección múltiple, particularmente en el contexto de su utilización para la evaluación de los resultados de las escuelas (Madaus y Kellaghan, 1992), señalan la necesidad de complementar su uso con otras formas de evaluación. Así, por ejemplo, además de las preguntas que piden al alumno que identifique una respuesta correcta entre una serie de posibles respuestas, se puede no facilitar ninguna opción al alumno de forma que tenga que ser él o ella misma quien la construya. Otros tipos de preguntas que requieren que el alumno elabore una respuesta incluyen aquéllas en las que éste ha de escribir una palabra, locución, frase, o incluso, un ensayo extenso en el que tenga que organizar sus ideas y presentarlas de un modo coherente y quizás persuasivo. También se ha utilizado una evaluación práctica más compleja (por ejemplo, pidiendo al alumno que prepare un experimento de ciencias).

En los últimos años se ha hablado mucho de la necesidad de lo que se llama evaluación de "desempeño", la cual implica tareas en las que se requiere que el alumno demuestre conocimientos y habilidades y elabore respuestas a tareas complejas. Estas tareas se utilizan para evaluar la competencia en áreas tales como las habilidades prácticas de medición en matemáticas o la realización de experimentos científicos. Idealmente, una evaluación de desempeño debería proporcionar información sobre los procedimientos que utilizan los alumnos, su aptitud para usar utensilios y la calidad de un producto terminado.

Mientras que las tareas de desempeño pueden tener bastantes ventajas—por ejemplo, las de facilitar modelos claros de resultados aceptables, influir positivamente en el aprendizaje y la instrucción y promover la enseñanza y el aprendizaje de formas más superiores de funcionamiento mental— presentan graves problemas por la posibilidad de variación en los procedimientos de administración y calificación de las mismas[3]. El enfoque más completo para el uso de exámenes de desempeño en una evaluación nacional se encuentra en el sistema de evaluación nacional británico. La clave del sistema, tal como fue concebido originalmente, eran las Tareas de Logro Estándar (Standard Attainment Tasks, SAT), las cuales fueron diseñadas para proporcionar información sobre el rendimiento de los alumnos en un grupo de metas de logros fijadas para una variedad de áreas del currículo. Las tareas fueron diseñadas para utilizar una gran variedad de modos de presentación (por ejemplo, oral, escrito, gráfico, video), de operación (solamente mental, escrito, oral práctico) y de respuesta (por ejemplo, elegir una opción dentro de una pregunta de selección múltiple, escribir una respuesta breve, escribir una composición o un producto práctico). Se exigió a los maestros que integraran las tareas en su práctica habitual en el aula, evitando de este modo la separación

artificial entre evaluación y enseñanza. También calificaban el desempeño de los alumnos.

La experiencia de la primera evaluación importante de niños de siete años en 1991 puso de manifiesto severas contradicciones en la administración y calificación de las SAT (Madaus y Kellaghan, 1992). La falta de estandardización que caracterizaba la administración y calificación, debe poner en tela de juicio el uso de los datos obtenidos para comparar los resultados individuales de un alumno o los resultados agregados de una escuela. Como consecuencia de los problemas sufridos en la administración de las SATs relativos al efecto del tema, al del calificador y a la generalización de los resultados, se han modificado enormemente los procedimientos de evaluación y en el futuro incluirán pruebas más simples y convencionales (Kellaghan, 1996b).

Otras variables

Los cuestionarios y las listas de clasificación por rango diseñados para obtener información contextual y relevante para la política se pueden administrar al mismo tiempo que los instrumentos de evaluación con un costo adicional relativamente pequeño. La información contextual puede incluir aquélla sobre los maestros (por ejemplo, sus calificaciones y frecuencia con la que asisten a cursos), el tamaño de la clase, la duración de la jornada escolar, el tiempo de enseñanza, las instalaciones de la escuela—como el número y la condición en que se encuentran los pupitres y libros—, la cantidad del libro de texto que se ha estudiado durante el año académico, el tiempo dedicado a las áreas del currículo, la cantidad de deberes que se mandan, el porcentaje de alumnos que tienen profesores fuera de la escuela y las tasas de asistencia, terminación y promoción de los alumnos.

La identificación de los factores contextuales relacionados con el rendimiento escolar puede ser útil especialmente para quienes toman las decisiones, los cuales pueden usar esta información para influir en la distribución de los escasos recursos financieros. El conocimiento de las variables contextuales puede anticiparse a las tendencias de los políticos a centrarse en una sola variable sin tener en cuenta otros posibles factores que pueden explicar un resultado. También puede ayudar a la identificación variables manipulables—por ejemplo, el tiempo dedicado a ciertas áreas del currículo, la naturaleza de la capacitación previa y en servicio del profesorado y las tasas de promoción de los alumnos—que parecen estar relacionadas de manera positiva con el rendimiento de ellos.

ADMINISTRACIÓN DE LA EVALUACIÓN NACIONAL

La logística de la administración de una evaluación nacional es compleja. Se tiene que contactar a las escuelas meta para asegurar su cooperación; imprimir, empaquetar y distribuir los materiales; reclutar y preparar al personal que administrará la evaluación; organizar visitas de supervisión a los centros de evaluación y recoger de las escuelas, limpiar, calificar y emparejar las hojas de

respuestas y los cuestionarios. Dada la cantidad de datos que genera una evaluación nacional, es conveniente crear al principio una base de datos con información sobre los alumnos, a la cual se añadirá más información cuando se reciban los materiales de las escuelas.

Existen dos maneras de abordar la administración de las tareas de evaluación en las escuelas. Una consiste en encargar a los maestros de la administración de las tareas en sus propias escuelas; la otra en utilizar personal externo que visite las escuelas.

Confiar lo más posible de la administración misma de las pruebas a los maestros en las escuelas en las que se llevan a cabo las evaluaciones reducirá sustancialmente los costos administrativos. La participación de los maestros también puede contribuir a la viabilidad política de la evaluación y a aumentar la probabilidad de que se llevarán a la práctica las reformas que surjan de la evaluación. Sin embargo, es posible que no todos los maestros sigan los procedimientos administrativos adecuadamente, lo cual originará problemas a la hora de realizar comparaciones.

Cuando exista una fundada preocupación de que pueda estar en juego la validez de la evaluación si se encarga la administración de las pruebas a los maestros, se deben adoptar estrategias alternativas. La utilización del inspectorado o del personal del currículo del Ministerio de Educación o los investigadores para administrar las pruebas y otros instrumentos puede ser una alternativa viable y es más probable que garantice que se sigan los procedimientos estándares en todas las clases. Aún más, la participación del personal del Ministerio puede otorgar cierta categoría al ejercicio. No obstante, también existen desventajas asociadas con este procedimiento. Primero, está el asunto del costo del tiempo de los vigilantes y, quizás, de su traslado y mantenimiento. Segundo, puede molestar a maestros y alumnos; y, tercero, está el tema de la administración del tiempo. Dado que las evaluaciones deben llevarse a cabo en todas las escuelas aproximadamente al mismo tiempo, se exigiría que un gran número de evaluadores externos cubrieran todo el país en un período limitado de tiempo.

Sea quien sea el responsable de gestionar los procedimientos de evaluación, es importante que las condiciones en las que se realiza la evaluación sean lo más uniformes posibles de una escuela a otra. Además de asegurar que todos los alumnos reciben las mismas o sus tareas equivalentes, debe haber igualmente uniformidad en las instrucciones, los materiales, el tiempo permitido para completar las tareas y el ambiente general de la evaluación.

Hay varias formas de resolver el problema de la variación de procedimientos. Primero, hacer las tareas de evaluación inequívocas y relativamente simples. Segundo, facilitar entrenamiento sobre la administración de las tareas. Tercero, proveer de un manual detallado a aquéllos que se responsabilizarán de la administración. El manual debe ser claro respecto al número de alumnos que se va a evaluar y al método a seguir para seleccionarlos si es que no han sido previamente seleccionados por la institución ejecutora. Las instrucciones deben incluir una

agenda de trabajo así como detalles precisos para la administración de los instrumentos y las tareas. Debe existir un buen suministro de materiales. Finalmente, es recomendable algún tipo de control de calidad de la administración de una evaluación. Esto supondría la participación de varias personas que estén completamente familiarizadas con los procedimientos exigidos, las cuales visitarían una muestra de escuelas durante la administración de la prueba para asegurarse de que se sigan los procedimientos. Obviamente, esta tarea debe realizarse con gran delicadeza y, aun así, puede que se molesten los maestros.

TIEMPO Y FRECUENCIA DE LA EVALUACIÓN

La frecuencia con la que se realiza una evaluación nacional varía de un país a otro. En Francia, todos los alumnos de tercero, sexto y décimo son evaluados cada año y una muestra de noveno es evaluada cada cinco años aproximadamente. En Inglaterra y Gales también todos los alumnos de siete, once y catorce años de edad son evaluados todos los años. En otros lugares, las evaluaciones son menos frecuentes. Por ejemplo, el Canadá sigue un ciclo de tres años y Finlandia, uno de diez. En algunos países no existe un ciclo predeterminado y las evaluaciones se llevan a cabo cuando se estima necesario.

A menos que exista un proceso de retroinformación a escuelas individuales (como sucede en Inglaterra y Gales) o a alumnos individuales (como en Francia), no parece necesario realizar una evaluación nacional. Si la finalidad de la evaluación es proporcionar información sobre el desempeño del sistema en su conjunto, lo adecuado sería realizar una evaluación de un área específica del currículo cada tres o cinco años. Los sistemas educativos no cambian rápidamente y lo más probable es que evaluaciones más frecuentes no mostrarán cambios.

Aunque se puede realizar una evaluación de un área del currículo cada cinco años solamente, esto no significa que no se puedan llevar a cabo evaluaciones de otras áreas en los años intermedios. Un sistema de evaluación nacional en el que, por ejemplo, se evalúan áreas básicas del currículo (lectura, matemáticas, ciencias) cada tres años puede incluir la evaluación de las tres áreas al mismo tiempo cada tres años o la de un área del currículo cada año.

El momento del año en que se realiza la evaluación estará en cierta medida determinado por la finalidad de la evaluación en su conjunto. Si, por ejemplo, el fin es obtener información sobre los rendimientos escolares durante el último año en que la mayoría de los alumnos comparten el mismo currículo (como en el caso del Canadá cuando se evalúa a los alumnos de trece años) o en el último año de enseñanza obligatoria (que es a los dieciséis años en la mayor parte de las provincias canadienses), entonces la evaluación se realizará hacia el final del curso escolar correspondiente.

Si, por el contrario, se van a utilizar los resultados de una evaluación nacional con fines de diagnóstico y se espera que influyan directamente sobre la enseñanza en

algunas clases en concreto, entonces se evaluará a los alumnos al principio del año académico. Este es el caso de Francia, donde se evalúa a los alumnos al entrar a la escuela secundaria inferior (sexto curso) y al principio de la educación secundaria superior (décimo curso). Es posible que un país realice evaluaciones en momentos diferentes del año, dependiendo del tipo de información que sea necesario. De este modo, en Francia, además de las evaluaciones de primeros de año, se realiza una al final del año en el último curso de la escuela secundaria inferior (noveno) para proporcionar información sumativa sobre cómo funciona el sistema en ese momento.

ANÁLISIS DE LOS DATOS

Cuando se terminan las tareas de evaluación en las escuelas, todos los materiales relacionados con ellas deben devolverse a la institución ejecutora. Debe existir un sistema de gestión de datos para recibir los materiales y comprobar que no falte nada. Puede que sea necesario ponerse en contacto con las escuelas que no han enviado todos sus materiales para averiguar las razones del retraso y, quizás, para animarles a llevar a cabo la evaluación si todavía no lo han hecho.

Cuando se retornen todos los materiales, se calificarán y registrarán las respuestas de los alumnos, se limpiarán los datos y se introducirán en una base de datos, que puede haber sido creada al principio del estudio o, si no, habrá de establecerse en este momento.

La forma de análisis a realizar dependerá de la decisión que se haya tomado sobre cómo se comunicarán los resultados (ver el próximo apartado). Cuando se han recopilado datos aparte de los relativos al rendimiento escolar, se deben diseñar análisis para identificar las relaciones entre el rendimiento del alumno y las variables de contexto u otras que no representan el rendimiento, las cuales incluirán probablemente características personales de los alumnos y las instalaciones de la escuela. Los resultados de tales análisis pueden ayudar a evitar que la gente llegue a conclusiones simplistas—por ejemplo, la de que las escuelas privadas son "mejores" que las públicas cuando la diferencia puede explicarse en gran medida por las distintas condiciones de los hogares de los alumnos.

Es probable que las comparaciones de los resultados de la evaluación de escuelas de distintos tipos, regiones o grupos (grupos étnicos, por ejemplo) provoquen gran interés en los políticos. Ciertamente, las diferencias entre los niños y las niñas o entre los alumnos de escuelas urbanas y los de las rurales deberían ser motivo de preocupación para políticos y directores de la educación. En efecto, uno de los propósitos de una evaluación nacional puede ser el obtener datos empíricos sobre el alcance de tales diferencias. No obstante, esos resultados deben ser sometidos a un profundo examen y no se deben interpretar de una manera simplista que implique causalidad cuando ésta no está justificada. Además, sería ingenuo considerar el mero reconocimiento (y posible publicación) de las diferencias como una respuesta adecuada a los problemas subyacentes en un desempeño diferencial.

PRESENTACIÓN DE LOS RESULTADOS

Los resultados de una evaluación deben comunicarse tan pronto como sea posible después de la recolección de los datos. Si se retrasan más de siete u ocho meses, la utilidad del ejercicio disminuye. Los informes deben ser concisos, claros y desprovistos de términos de la jerga educativa. Pueden servir como modelos los oportunos, bien presentados y bien ilustrados informes producidos por la NAEP y, recientemente, por el TIMSS. Se debe considerar la publicación de informes resumidos para el público en general y otros más detallados para políticos, gestores educativos y maestros.

Se han utilizado muchas maneras de presentar los resultados en evaluaciones nacionales e internacionales. Una consiste en presentar los niveles medios de rendimiento escolar en un área del currículo. Las otras implican informar sobre el porcentaje de alumnos asociados con rendimientos específicos. Los rendimientos, por su parte, se definen de maneras diversas.

Desempeño promedio de los estudiantes en un área del currículo

Si se suman los puntajes individuales de una muestra representativa de alumnos en un país y después de divide entre el número de alumnos, se obtiene el promedio total del rendimiento en un área particular del currículo, de una edad o curso en concreto, en ese país. En la práctica, pueda que el procedimiento no sea tan simple como esto, ya que quizá haya que realizar ajustes para reflejar el muestreo desproporcionado de alumnos en diferentes regiones geográficas o tipos de escuelas. Lo esencial es que se está tratando de reflejar el nivel medio de desempeño en el país en términos cuantitativos.

Esta información tiene un valor limitado por varias razones. Primero, no nos dice si el promedio obtenido puede ser considerado "satisfactorio" o "insatisfactorio"—a menos que existan datos comparativos con los cuales se pueda contrastar dicho resultado promedio. Así, por ejemplo, se podría utilizar la información como una indicación de si los estándares se mantienen, en general, estables en el país, o si suben o bajan , si existiese información de un momento anterior en el tiempo con la que pudiera compararse. También sería útil si existiese información similar de otros países, como sucede en los estudios internacionales de evaluación. Tanto la IEA como la IAEP han comunicado las puntuaciones medias alcanzadas por los países participantes en una variedad de áreas del currículo. La Organización para la Cooperación y el Desarrollo Económico (OCDE) ha utilizado estos datos para subrayar diferencias en el rendimiento entre sus países miembros (OCDE, 1995).

Un segundo problema que se deriva del uso de puntuaciones agregadas medias para un área amplia del currículo como las matemáticas o las ciencias, consiste en que, al sumar el número de preguntas acertadas, se están ocultando las diferencias de desempeño existentes entre distintos aspectos de una asignatura, tales como, en el

caso de las matemáticas, entre la aptitud de cálculo y la aptitud de resolución de problemas. Es decir, conviene señalar las puntuaciones medias por campos, en qué puntos del currículo se encuentran concentradas de una forma significativa, ya que las diferencias entre los desempeños en los distintos campos puede tener valor de diagnóstico.

También puede interesar conocer los promedios de sectores del sistema educativo. En la evaluación nacional colombiana, por ejemplo, se calcularon los promedios de rendimiento para cada área del currículo por estado, situación (urbana/rural) y tipo de institución (pública/privada) (Rojas, 1996).

Porcentaje de respuestas acertadas

Algunas evaluaciones nacionales (por ejemplo, la NAEP en los Estados Unidos) e internacionales (como la Evaluación Internacional del Progreso Educativo, International Assessment of Educational Progress, IAEP) han comunicado resultados a nivel de pregunta individual. Entonces, se utilizaron estadísticas del porcentaje de respuestas correctas promedio para resumir los resultados[4]. Esta forma de comunicación es válida, aunque probablemente sea demasiado detallada para la mayor parte de los lectores. Es más, si se tienen que hacer comparaciones entre evaluaciones o entre los resultados para cursos diferentes, este enfoque exige que se usen grupos idénticos de preguntas.

Porcentaje de alumnos que logran dominar los objetivos del currículo

Otro enfoque es el que presenta los porcentajes de alumnos que logran dominar los principales objetivos del currículo. Este es un planteamiento intermedio entre la comunicación de estadísticas por pregunta y la de estadísticas de campos amplios dentro de un área del currículo. En una evaluación irlandesa, se dividió el currículo de matemáticas para los alumnos de los cursos quinto y sexto en 55 objetivos en cálculo, conceptos y resolución de problemas. Por objetivos se entendía, por ejemplo, el que un alumno fuera capaz de: sumar una columna de números de no más de cinco dígitos, restar dos números de no más de tres dígitos, llevar a cabo operaciones simples de aritmética en las que formaba parte el cero, e identificar factores comunes entre dos números. Se consideraba que un alumno había dominado un objetivo cuando él o ella había respondido un número determinado de preguntas de cada objetivo dentro de una prueba escrita de selección múltiple. Se proporcionaron estadísticas para cada uno de los 55 objetivos, indicando el porcentaje de alumnos que habían dominado el objetivo. Se reseñaron también los aspectos del currículo nacional que habían presentado problemas (Kellaghan, Madaus, Airasian y Fontes, 1976).

Porcentaje que logra metas específicas de conocimientos

En algunos sistemas educativos se establecen, para los alumnos, metas específicas de conocimientos en varios momentos de sus carreras educativas. Donde sucede esto,

se puede diseñar un sistema de evaluación con el fin de obtener un cálculo aproximado del número de alumnos que están alcanzando esas metas. En el sistema británico, se señala la medida en la que los alumnos están logrando las metas de conocimientos del currículo nacional para los siete, once, catorce y dieciséis años de edad. Cada meta está dividida en niveles de creciente dificultad en una escala de uno a diez, con criterios claros para definir lo que un alumno debe saber, entender o ser capaz de hacer para que se considere que su puntuación se sitúa en un determinado nivel. En la evaluación de 1991, había 32 metas relevantes para los niños de siete años: cinco en inglés, 13 en matemáticas y 14 en ciencias. Los resultados se presentaron como porcentajes de alumnos que cumplían con los requisitos de cada nivel en cada área del currículo.

En inglés, se indicaron los porcentajes de quienes alcanzaban los niveles uno, dos y tres para cinco metas: comunicación oral y escrita, lectura, escritura con un sentido, ortografía y escritura a mano. En matemáticas, ejemplos de las metas para las cuales se presentaron resultados fueron número, álgebra y medidas, uso y utilización de las matemáticas, números y notación numérica, operaciones con números (+, -, ÷, x) y forma y espacio (formas bi- y tri-dimensionales). Las metas de ciencias incluían procesos de vida, genética y evolución, influencias humanas sobre la tierra, tipos y usos de materiales, energía y sonido y música (Great Britain, Department of Education and Science, 1991).

Porcentaje que funciona a determinados niveles de competencia

Otro modo de presentar los resultados utilizado en varias evaluaciones estatales y nacionales consiste en elaborar una escala de competencia por procedimientos estadísticos y determinar niveles dentro de la escala a través de procesos de juicio. Se han fabricado escalas de competencia tanto para evaluaciones nacionales (el Canadá, los Estados Unidos) como para las internacionales (IEA, IAEP). Por ejemplo, en la evaluación nacional canadiense de niños de trece y dieciséis años, se establecieron cinco niveles de competencia en matemáticas, ciencias, lectura y escritura (Canada, Council of Ministers of Education, 1996). Cada nivel es descrito en términos del conocimiento y las aptitudes que debe mostrar un alumno que operase a ese nivel; se presenta, entonces, el porcentaje de alumnos que funciona a cada nivel. En ciencias, el alumno debería ser capaz de describir, en un nivel determinado:

- Nivel 1. Propiedades físicas de los objetos.
- Nivel 2. Cambios cualitativos en las propiedades de una sustancia cuando ésta se calienta o se enfría.
- Nivel 3. Estructura de la materia en términos de partículas.
- Nivel 4. Cualitativamente, una reacción química o cambio de fase.
- Nivel 5. Cuantitativamente, el producto de una reacción dadas las sustancias, o vice versa.

A veces se añaden etiquetas a los niveles. Por ejemplo, se describe a los alumnos de la NAEP como carentes de la competencia básica, que han alcanzado la

competencia básica, competentes, o avanzados. Una nomenclatura alternativa se utilizó en una evaluación en Kentucky: se describió a los alumnos como novicios, aprendices, expertos, o distinguidos (Guskey, 1994). Aunque tales etiquetas tienen una obvia atracción intuitiva, pueden tener connotaciones negativas. También pueden significar cosas diferentes para distintos niveles e incluso, en el mismo curso, es probable que sean interpretadas de formas diversas por personas diferentes. Cuando se presentan los resultados por niveles, sería preferible evitar las etiquetas, utilizando en su lugar descripciones verbales de lo que un alumno en un nivel sabe o es capaz de hacer.

Puede que existan problemas más graves asociados con el establecimiento de escalas. Uno es el supuesto de que los resultados de los estudiantes en una evaluación están determinados por el valor de un "rasgo" exclusivamente y que, en consecuencia, el conjunto de preguntas se pueden considerar unidimensionales, reflejando la "aptitud de lectura", la "aptitud matemática" o la "aptitud científica". Existen razones, surgidas de una consideración de la naturaleza del rendimiento escolar y de los procedimientos estadísticos implicados, que llevan a cuestionar la validez de este supuesto (Goldstein, 1996).

Los efectos prácticos del establecimiento de escalas sobre la representación y el equilibrio del currículo se pueden observar en los intentos por establecer niveles de rendimiento (básico, competente y avanzado) para la prueba de matemáticas NAEP de 1991 en los Estados Unidos para los cursos cuarto, octavo y décimo segundo. Al considerar estos niveles, el Órgano de Gobierno de la Evaluación Nacional (National Assessment Governing Body) se sorprendió por la insuficiencia del banco de preguntas y, en particular, por la falta de lo que ellos denominan preguntas "suficientemente estimulantes". Lo que había sucedido es que se habían excluido del conjunto de preguntas las que eran más difíciles y no cumplían con los criterios de la escala, ya que no contribuían a la escala sino que reducían la congruencia entre el procedimiento de evaluación y el campo del currículo para representar para el cual habían sido diseñadas (Kellaghan y Grisay, 1995). La pregunta a abordar en este contexto es: ¿Se debe aceptar una cobertura diferente del currículo con el fin de cumplir con los estándares técnicos o se debe mantener la postura de que lo más importante es la cobertura del mismo?

Distribución del rendimiento

Además de la información sobre el rendimiento medio, es útil conocer la distribución de los rendimientos de los alumnos en un área del currículo. Tal información se centra en las disparidades entre los alumnos que logran altos resultados y los que los muestran bajos en el sistema educativo. El uso de escalas de competencia, si se admite su validez, logra hacer esto. Pero, incluso si no se utilizan tales escalas, un examen de los datos en bruto junto con un proceso de juicio sobre la suficiencia de los desempeños de los alumnos, puede proporcionar ideas útiles. Las disparidades en el rendimiento se pueden analizar por sexo, situación geográfica o tipo de escuela para facilitar una lectura más rica de los datos arrojados por la

evaluación nacional. Puede suceder, por ejemplo, que se encuentre una mayor incidencia de alumnos con calificaciones bajas en algunas áreas del país que en otras y entre niños que entre niñas.

CÁLCULO DE LOS COSTOS

No es posible calcular los costos definitivos de una evaluación nacional que serían aplicables a todas las situaciones ya que éstas variarán dependiendo del nivel de desarrollo socioeconómico del país y las características de la evaluación que se está llevando a cabo. Se esperarían costos más altos para sueldos y servicios en un país industrializado que en uno en vías de desarrollo, aunque estos últimos quizás tengan que importar la pericia necesaria, lo cual puede, desde luego, ser costoso.

La información disponible corrobora la idea de que el costo de las evaluaciones nacionales varía considerablemente. Así, en Jamaica, se ha estimado que el costo por alumno evaluado es de un dólar americano para 20.000 alumnos presentados al examen y de 1,32 dólares americanos para 10.000 examinandos (Ilon, 1996). El costo estimado del Programa de Evaluación del Rendimiento Escolar (PER) chileno para los años 1981-84 fue de cinco dólares americanos por alumno (Himmel, 1996). Para el Programa Nacional de Medición de la Calidad de la Educación Básica Chilena (SIMCE) de 1988, se calcula el costo en 4,31 dólares americanos por alumno en el cuarto curso y 6,94 dólares americanos por alumno en el octavo curso (Olivares, 1996).

Dado que las condiciones de una evaluación y las circunstancias en las que se realiza van a influir en el costo, es necesario calcular el costo de cada evaluación individual antes de llevarla a cabo. Ilon (1996) ha señalado las consideraciones que han de tenerse en cuenta en este ejercicio, que es probable que incluya a especialistas en evaluación, políticos, educadores y un economista. Estas personas tendrán que colaborar en un proceso interactivo en el que se modifiquen los componentes propuestos de la evaluación hasta que los costos encajen dentro de los márgenes del presupuesto disponible. Quienes participen en el proceso se darán cuenta en seguida que cada componente de la evaluación implica decisiones que sopesan los costos y la información que los interesados querrían obtener. Quizás haya que ceder en lo que se desea, a la luz de las limitaciones presupuestarias.

Ilon plantea una serie de preguntas sobre los diversos componentes de una evaluación nacional que han de tenerse en cuenta cuando se están considerando los costos.

Comité promotor. Es probable que el costo de un comité de dirección sea pequeño comparado con los gastos totales. Sin embargo, los costos pueden variar dependiendo del alcance del transporte, la administración requerida y si los miembros son o no compensados por su tiempo. Los costos de transporte dependen, obviamente, de qué tan dispersos se encuentren los números mientras que el tema de la compensación dependerá de la costumbre.

Institución ejecutora. Los costos variarán dependiendo de si la institución cuenta con las instalaciones, la experiencia y la pericia necesarias para realizar la evaluación nacional o si tendrá que mejorar sus instalaciones, proporcionar formación adicional para el personal y/o contratar consultores externos, cada uno de los cuales pueden suponer un gasto considerable.

Obtención de apoyo. Puede suceder que el apoyo para una evaluación nacional se encuentre relativamente extendido en un país, en cuyo caso bastará con un simple paquete de información. Sin embargo, si es necesario aumentar la conciencia sobre el valor de una evaluación nacional y conseguir el apoyo de maestros, padres y del público en general, se puede incurrir en gastos considerables tanto en contratar gente para producir folletos, videos y carteles, como en transmitir el mensaje a través de reuniones, charlas, radio y televisión.

Población meta. Varios factores relativos a la elección de la población meta tienen repercusiones en el costo. Primero, el número de cursos o grupos de edades que se van a evaluar es relevante no sólo para la administración de un procedimiento de evaluación sino también para el costo de la preparación de instrumentos. Segundo, los costos se verán afectados tanto en la administración como en la producción de materiales si se decide examinar a toda la población de una determinada edad o curso en vez de a una muestra. Tercero, seleccionar una población que abarca diferentes tipos de escuelas aumentará considerablemente el costo. Por ejemplo, si la población meta es niños de trece años de edad y encontramos estudiantes de esta edad tanto en las escuelas de primaria como en las de secundaria, será necesario hacer un muestreo de un número suficiente de escuelas de primaria y secundaria para obtener una imagen precisa de los rendimientos de la población estudiada. Cuarto, seleccionar un grupo de edad será probablemente más caro que elegir un curso ya que los alumnos de una edad determinada pueden estar repartidos por una serie de cursos, lo cual puede requerir, en ocasiones, sesiones adicionales de exámenes. No obstante, es más probable que sea la escuela y no la organización ejecutora de la evaluación nacional, la que corra con este gasto, a menos que esta última esté pagando por la vigilancia.

Contenido y elaboración del instrumento. Las diversas opciones para la selección del contenido y la forma de la evaluación deben ser consideradas en términos de costo así como en otros términos, tales como la validez y la facilidad de administración. La elaboración de los procedimientos de evaluación, su administración y calificación tienen incidencia en los costos. Los exámenes de selección múltiple son relativamente caros de elaborar pero relativamente baratos de calificar. Lo mismo se puede decir de las evaluaciones de desempeño que conllevan el uso de una cinta o un video y en las cuales se pide al alumno que grabe su respuesta. En el caso de estas técnicas, sin embargo, puede que haya costos en términos de equipo (por ejemplo, la grabadora) que las escuelas quizás no tengan y que sea necesario transportarlo. Otras formas de evaluación del desempeño en las que se pide a los alumnos que escriban un ensayo largo o que demuestren algún

conocimiento o destreza prácticos en presencia de un evaluador (por ejemplo, realizar un experimento de ciencias, tocar un instrumento musical) puede que no sean tan caras de diseñar pero son más caras de aplicar y calificar.

Manuales de administración. utilizados para asegurar uniformidad en la aplicación de una evaluación y pueden variar en su forma de presentación, la calidad del papel, el uso del color y la encuadernación pero, en general, la preparación del manual no debería acarrear grandes costos. No se debe olvidar el costo de distribución.

La administración o recolección de datos será el elemento más caro de una evaluación nacional. Supone una serie de pasos. Antes de la evaluación, se puede obtener de las escuelas información relativa a sus alumnos elegibles, de entre los cuales la institución ejecutora seleccionará a los que participarán en la evaluación. Se han de diseñar e imprimir los materiales de los exámenes y, a continuación, se deben clasificar, empaquetar y enviar a las escuelas. También se tienen que tomar las medidas necesarias para la devolución de los materiales y para cubrir los gastos que ello implica. Un factor muy importante que se debe tener en cuenta en la recolección de datos es si se requieren vigilantes especiales que visiten las escuelas o si serán los maestros quienes apliquen las pruebas en sus propias escuelas. En cualquier caso, pero sobre todo en el último, se tendrá que atender a la supervisión de la administración de las tareas de evaluación.

Análisis. Los costos analíticos dependerán del tipo de procedimiento de evaluación utilizado y de la disponibilidad de tecnología para llevar a cabo la calificación y el análisis. Mientras normalmente se considera que la calificación de pruebas de selección múltiple por medio de máquinas es mucho más barata que la calificación manual, puede que éste no sea el caso en un país en donde los costos tecnológicos son elevados y la mano de obra es barata. El costo del análisis que producirá la información sobre los niveles de rendimiento de los alumnos dependerá de cómo se decida presentar los resultados. El método más caro supone calificar e informar en términos de niveles de competencia.

Presentación de los resultados. Se deben apartar fondos para escribir un informe y para el caso de que haya que hacer cambios en el mismo para el gobierno, los maestros y el público en general.

Componentes del costo. Como señalamos antes, es probable que los costos de una evaluación nacional varíen de un país a otro. Aunque las proporciones del costo relativas a los diversos componentes pueden también cambiar, las proporciones utilizadas en la NAEP de los Estados Unidos pueden servir de guía aproximada para un ejercicio de cálculo del costo. La recolección de los datos (30 por ciento del costo total) fue el elemento más caro, seguido de la elaboración del instrumento, el análisis de los datos y la preparación y distribución del informe, cada uno de los cuales consumió 15 por ciento de los fondos. El muestreo y la selección (10 por ciento), procesamiento de los datos (10 por ciento) y gobierno (5 por ciento) fueron los elementos menos caros (Koeffler, 1991).

El consejo de Loxley (1992) de apartar unos cuantos fondos para emergencias, si bien dirigido a aquéllos que participan en evaluaciones internacionales, también es relevante para evaluaciones nacionales. Al recomendar que 10 por ciento del presupuesto se reserve para este fin, señala que "no se trata de si surgirán o no emergencias, sino de cúando y cuántas" (Pág. 293).

CONCLUSIÓN

Poca gente dudaría del valor de tener información sistemática sobre lo que los alumnos aprenden en la escuela. Tal información tiene obvias ventajas intuitivas, especialmente si se dispone de ella en diferentes momentos en el tiempo, lo cual permite emitir un juicio sobre la estabilidad y el ascenso o descenso de los estándares. No obstante, para lograr tal información es necesario iniciar un ejercicio de recolección de datos que cuesta dinero y que requiere muchas decisiones prudentes para asegurar que se obtenga la información que parece más adecuada desde el punto de vista de la política y la gerencia del sistema escolar.

Mientras se puede aprender mucho de la experiencia de otros países que han realizado evaluaciones nacionales, se debería admitir que esa experiencia muestra una variación considerable. Por ejemplo, al decidir sobre la institución que llevará a cabo la evaluación, mucho dependerá de las costumbres del país, mientras la información en concreto que se obtiene para los indicadores dependerá de la experiencia del país en el desarrollo de pruebas así como en las necesidades de quienes toman las decisiones. Los países que cuentan con una larga tradición de investigación educativa pueden encontrarse con que una evaluación nacional encaja con bastante facilidad en las estructuras existentes. Los países que no tienen esa tradición y que se están planteando emprender una evaluación nacional, por otro lado, tendrán que pararse a pensar muy seriamente los pasos que acarrea un ejercicio de tal índole y cómo van ellos a conseguir los recursos necesarios para llevarla a buen término.

Para un país que se está planteando realizar una evaluación nacional y que no tiene una infraestructura desarrollada para llevar a cabo una investigación educativa que incluye la administración de exámenes a gran escala, puede merecer la pena señalar que una evaluación nacional no debe ser demasiado ambiciosa en las áreas del currículo que cubre, los procedimientos de evaluación, la complejidad de la muestra o las demandas al personal. Casi inevitablemente existirá una tensión entre el ideal de recopilar cuanta más información sea posible y la necesidad de utilizar el ejercicio inicial para proporcionar datos básicos sobre el funcionamiento del sistema educativo y para desarrollar la capacidad local. Mantener manejable el alcance de una evaluación—por ejemplo, limitándola a un área del currículo y a un curso—aumenta la probabilidad de que la operación tenga éxito. Otra opción consiste en limitar su cobertura geográfica. Especialmente en países grandes y diversos, se puede obtener valiosa experiencia e información útil relacionadas con la política a través de evaluaciones limitadas a una o unas pocas regiones de un país. No ob-

stante, si se espera usar los datos de las evaluaciones nacionales para realizar un seguimiento de las tendencias de rendimiento a lo largo del tiempo, las limitaciones en los procedimientos de recolección de datos establecidas al comienzo afectarán la capacidad para realizar comparaciones en años posteriores.

El primer paso para cualquier país en proceso de decidir si realiza o no una evaluación nacional es determinar su costo. Solamente cuando se sabe en qué gastos se va a incurrir se puede empezar a plantear si sería mejor emplear el dinero requerido en alguna otra actividad educativa, tal como la mejora de las instalaciones escolares o la formación del profesorado.

Al decidir si se lleva a cabo una evaluación nacional, también es importante tratar de determinar qué tipo de información de tal evaluación es probable que sea utilizada por quienes toman las decisiones, así como considerar la forma en la que éstos podrán intervenir en el sistema educativo a la luz de la información revelada por la evaluación. En este contexto, se puede reflexionar sobre algunas de la preguntas siguientes a las cuales quizás responda una evaluación nacional.

- En el sistema educativo, ¿es el rendimiento, en conjunto, de los alumnos en un área precisa del currículo (por ejemplo, lectura, matemáticas, ciencias), en un curso o edad determinado peor o mejor de lo esperado?
- ¿ Es el desempeño de los estudiantes en campos particulares del rendimiento (por ejemplo, la comprensión de conceptos matemáticos) mejor o peor que en otros campos (por ejemplo, aptitud para realizar cálculos matemáticos)?
- ¿Deseamos obtener datos de base que nos permitirán juzgar si el desempeño de los alumnos en áreas clave del currículo (por ejemplo, lectura, matemáticas) está mejorando o empeorando con el tiempo?
- ¿Es el desempeño de los alumnos en ciertos sectores del sistema educativo menos satisfactorio que el de los alumnos en otros sectores? Esta pregunta puede formularse respecto a:
    - alumnos de escuelas urbanas y los de escuelas rurales;
    - alumnos de escuelas públicas y los de escuelas privadas;
    - alumnos de diversos grupos étnicos/lingüísticos;
    - alumnos de áreas geográficas diferentes (provincias o estados).
- ¿Varían, por género, las aptitudes y los conocimientos?
- ¿Existe evidencia de una relación entre los conocimientos y aptitudes de los alumnos y los factores tales como el tipo de formación que han tenido los maestros, la cantidad de tiempo que éstos dedican a un área del currículo en la clase o la que los alumnos emplean en hacer los deberes?

Una evaluación nacional correctamente diseñada puede arrojar luz sobre estos asuntos. Normalmente no indicará, sin embargo, soluciones específicas a ninguno de los problemas que denunciará. En la búsqueda de soluciones, es probable que los políticos y gestores vuelvan frecuentemente a soluciones ya probadas, escogiendo la que parezca más adecuada. Éstas incluyen la mejora de la gestión

educativa, el aumento de la disponibilidad de libros de texto, la mejora de la supervisión de la enseñanza, la formación del profesorado (tanto previa como en servicio) y la reforma de los programas de estudios (por ejemplo, exigiendo que se dedique más tiempo a un área determinada del currículo).

Al considerar maneras de abordar la reforma educativa con base en los resultados de una evaluación nacional, conviene recordar que la opción de un administrador o gestor educativo al ordenar una acción para hacer frente a las deficiencias reveladas en la evaluación está en gran medida limitada por nuestra ignorancia sobre cómo funciona el sistema educativo. Aunque nuestro conocimiento sobre la enseñanza es sólo parcial y no considera una teoría integrada de organización, procesos y efectos escolares (Bryk y Hermanson, 1993), existe el peligro de que, dado que las evaluaciones nacionales están basadas en un modelo de educación de insumo-resultados, las recomendaciones para abordar los problemas se basarán en tal modelo, suponiendo que el mismo representa adecuadamente el proceso educativo. Tal enfoque se podría calificar de miope. Es importante que aquéllos que tienen la responsabilidad de mejorar el servicio educativo miren más allá de simples modelos de insumo-resultados e intenten ser más creativos en su búsqueda de soluciones, dándose cuenta de que, en el pasado, una comprensión insuficiente ha llevado con frecuencia a intentos externos (gobierno) por controlar los insumos de la escuela que no lograban generar las consecuencias pretendidas.

Al plantearse reformas basadas en los datos de una evaluación nacional, no se debería perder de vista el hecho de que los aspectos de la enseñanza que no están totalmente representados en el modelo de insumo-resultados, o que incluso se hallan mal representados, pueden jugar, de hecho, un papel importante influyendo en el rendimiento de los alumnos. Por ejemplo, si pensamos en las escuelas como sistemas, se hace aparente que el modelo de función de producción de la escolarización, caracterizado principalmente por causas y efectos independientes, no puede enfrentarse adecuadamente con el flujo de información y comunicación que conlleva una variedad de ciclos de retroinformación que es una característica de los sistemas (Bryk y Hermanson, 1993). Aún más, si es cierto que los aspectos comunitarios de la enseñanza tienen importantes repercusiones para el aprendizaje de los alumnos, entonces de nuevo el enfoque de insumos-resultados de nuestro sistema de indicadores actual no representará adecuadamente lo que sucede en las escuelas y, como consecuencia, se hallará limitado en su capacidad para diagnosticar problemas o actuar como base para una acción correctiva.

Estas consideraciones han llevado a una serie de comentaristas a sugerir la necesidad de utilizar tipos de indicadores diferentes de los que actualmente se usan y para un fin diverso según la información que proporcionan. Se ha propuesto, por ejemplo, que se enmarcara un sistema de indicadores en torno de seis grandes áreas temáticas que preocupan en educación: resultados del aprendizaje, calidad de las instituciones educativas, preparación de los niños para aprender cuando llegan a la escuela, apoyo social al aprendizaje, contribución de la educación a la productividad económica y equidad. Más aún, se ha llegado a sugerir que los datos derivados de los indicadores

que se relacionan con estas áreas se deberían encaminar a informar a diferentes públicos sobre el estado del sistema educativo, en vez de ser utilizados como base para una acción instrumental. En línea con la necesidad de informar, se debería trabajar para facilitar una comprensión a fondo de las fuerzas que se esconden bajo los indicadores clave. Para profundizar este entendimiento, se recomiendan investigaciones, estudios de caso y evaluaciones de programas (Bryk y Hermanson, 1993).

Es difícil distinguir estas actividades de las que han constituido los pilares de la investigación educativa durante mucho tiempo. Si el mensaje en el caso a favor de una forma diferente de indicador significa que el trabajo actual de seguimiento en el que se utilizan indicadores debe ser abandonado a favor de una investigación educativa más tradicional, es poco probable que se le preste atención. Primero, la investigación educativa tradicional ha contribuido de forma modesta a cubrir las necesidades de los administradores y gestores educativos. Aún más, el tipo de programa previsto como alternativa sería extremadamente caro. Muchos políticos consideran—correcta o equivocadamente— que el actual movimiento en torno de los indicadores proporciona información a quienes toman las decisiones que es más relevante que la información que arrojaba la investigación educativa tradicional y a un precio más asequible. No deben, sin embargo, olvidar las inquietudes que se esconden tras la búsqueda de indicadores alternativos y una forma alternativa de responder a la información del indicador. El sistema actual tiene cuando menos serias limitaciones y cuando más serios defectos que no deben pasarse por alto.

Sea cual sea la razón, la idea de una evaluación nacional y su correspondiente sistema de indicadores se está haciendo más popular a través del mundo y parece que será difícil para los países en un futuro inmediato resistirse a la tendencia a documentar los resultados de sus empresas educativas. Es de esperar, no obstante, que esto se llevará a cabo sensatamente y que no se realizarán evaluaciones nacionales simplemente por realizarlas. El tiempo usado en considerar seriamente el tipo de información que puede ser útil y con qué precisión se utilizará será tiempo bien empleado antes de embarcarse en una evaluación nacional.

NOTAS

[1] Véase, por ejemplo, Chinapah, V. (1996). After Jomtien: UNESCO's current policy on assessment. En A. Little y A. Wolf (Ed.) Assessment in transition. Learning, monitoring and selection in international perspective. Oxford: Pergamon. Págs. 42-57.

[2] Véase Weiss, C.H. (1979). The Many Meanings of Research Utilization. Public Administration Review, 39, 426-431.

[3] Véanse Gipps, C. y Murphy, P. (1994). A Fair Test? Assessment, Achievement and Equality. Buckingham, UK: Open University; Mehrens, W. A. (1992). Using Performance Assessment for Accountability Purposes. Education Measurement: Issues and Practice, 11

(1): 3-9, 20; Meisels, S. J., Dorfman, A. y Steele, D. (1995) Equity and Excellence in Group-administered and Performance-based Assessments. In M. T. Nettles y A. L. Nettles (Ed..), Equity and Excellence in Educational Testing and Assessment. Boston: Kluwer: 243-261.

[4] Véanse Baker, E.L. y Linn, R. L. (1995). United States. En: Performance Standards in Education: In Search of Quality. París: Organización para la Cooperación y el Desarrollo Económico: 197-209; Phillips, G. W., Mullis, V. S., Bourque, M. L., Williams, P. L., Hambleton, R. K., Owen, E. H. y Barton, P. E. (1993). Interpreting NAEP Scales. Washington, DC: National Center for Education Statistics.

REFERENCIAS

Baker, E.L., y Linn, R.L. (1995). "United States." En: *Performance Standards in Education: In Search of Quality*. París:Organización para la Cooperación y el Desarrollo Económico: 197-209.

Bloom, B.S., Madaus, G.F., y Hastings, T. (1981). *Evaluation to Improve Student Learning*. Nueva York: McGraw Hill.

Bottani, N., y Tuijnman, A. (1994). "International education indicators: Framework, development and interpretation." En: *Making Education Count. Developing And Using International Indicators*. París: Organización para la Cooperación y el Desarrollo Económico: 21-35.

Bryk, A. S., y Hermanson, K. L. (1993). "Educational indicator systems: Observations on their structure, interpretation, and use." *Review of Research in Education, 19:* 451-484.

Burnstein, L., Oakes, E., y Guiton, G. (1992). "Education indicators." En: M.C. Alkin (Ed.), *Encyclopedia of Educational Research* (6th ed). Nueva York: Macmillan: 409-418.

Canada. Council of Ministers of Education. (1996). *School Achievement Indicators Program. Science Assessment Framework and Criteria*. Toronto, Ontario: Autor.

Chinapah, V. (1992). *Monitoring and Surveying Learning Achievements: A status report*. Studies and Working Documents 1. París: UNESCO.

Chinapah, V. (1996). "After Jomtien: UNESCO's current policy on assessment." En: A. Little y A. Wolf (Ed.). *Assessment in Transition. Learning, Monitoring And Selection in International Perspective*. Oxford: Pergamon: 42-57.

Elley, W. B. (1992). *How In The World Do Students Read? IEA Study Of Reading Literacy*. La Haya: International Association for the Evaluation of Educational Achievement.

Frith, D. S., y Macintosh, H. G. (1984). *A Teacher's Guide To Assessment*. Cheltenham, UK: International Association for Educational Assessment.

Gipps, C., y Murphy, P. (1994). *A Fair Test? Assessment, Achievement And Equality*. Buckingham, UK: Open University.

Goldstein, H. (1996). "International comparisons of student achievement." En: A. Little y A. Wolf (Ed.). *Assessment In Transition. Learning, Monitoring And Selection In International Perspective*. Oxford: Pergamon: 58-87

Greaney, V., y Kellaghan, T. (1996). *Monitoring The Learning Outcomes Of Educa-*

tion Systems. Washington, DC: Banco Mundial.

Greaney, V., y Rojas, C. (1996). "Lessons learned." En: P. Murphy, V. Greaney, M. E. Lockheed y C. Rojas (Ed.). *National Assessments: Testing The System*. Washington, DC: Banco Mundial: 169-173

Great Britain. Department of Education and Science (1991). *Testing 7-Year-Olds In 1991: Results Of The National Assessment In England*. Londres: Autor.

Guskey, T. R. (1994). *High Stakes Performance Assessment: Perspectives On Kentucky's Educational Reform*. (2nd ed). Thousand Oaks, CA: Corwin Press.

Guthrie, J. W. (1991). "The world's new political economy is politicizing educational evaluation." *Educational Evaluation and Policy Analysis, 13:* 309-321.

Himmel, E. (1996). "National assessment in Chile." En: P. Murphy, V. Greaney, M.E. Lockheed, y C. Rojas (Ed.). *National Assessments: Testing The System*. Washington DC: Banco Mundial: 111-128.

Ilon, L. (1996). "Considerations for costing national assessments." En: P. Murphy, V. Greaney, M. E. Lockheed y C. Rojas (Ed.). *National Assessments: Testing The System*. Washington, DC: Banco Mundial: 69-88.

Johnson, E.G. (1992). "The design of the National Assessment of Educational Progress." *Journal of Educational Measurement, 29:* 95-110.

Kellaghan, T. (1996a). "IEA studies and educational policy." *Assessment In Education, 3*: 143-160.

Kellaghan, T. (1996b). "National assessment in England and Wales." En: P. Murphy, V. Greaney, M. E. Lockheed, y C. Rojas (Ed.). *National Assessments: Testing The System*. Washington, DC: Banco Mundial: 129-136.

Kellaghan, T., y Grisay, A. (1995). "International comparisons of student achievement: Problems and prospects." En: Organisation for Economic Co-operation and Development, *Measuring What Students Learn*. París: OCDE: 41-61.

Kellaghan, T., Madaus, G.F., Airasian, P. W., y Fontes, P. J. (1976). "The mathematical attainments of post-primary school entrants." *Irish Journal of Education, 10*: 3-17.

Koeffler, S. (1991). *Assessment Design*. Documento presentado en el Seminario sobre Medición/Evaluación, Educational Testing Service, Princeton, NJ.

Lapointe, A. (1990). *Why a national assessment?* Princeton, NJ: National Assessment of Educational Progress.

Lee, V.E., Bryk, A.S., y Smith, J.B. (1993). "The organization of effective secondary schools." *Review of Research in Education*, 19: 171-267.

Lockheed, M. E., y Hanushek, E. (1988). "Improving educational efficiency in developing countries: What do we know?" *Compare, 18* (1): 21-38.

Loxley, W. (1992). "Managing international survey research." *Prospects, 22:* 289-296.

Madaus, G.F., y Kellaghan, T. (1992). "Curriculum evaluation and assessment." En: P. W. Jackson (Ed). *Handbook Of Research On Curriculum*. Nueva York: Macmillan: 119-154.

Martin, M. O., y Kelly, D. L. (1996). *Third International Mathematics and Science Study. Technical report. Vol 1: Design and development*. Chestnut Hill, MA: Boston College.

Martin, M. O., Mullis, I. V. S., Beaton, A. E., Gonzalez, E. V., Smith, T. A., y Kelly, D. L. (1997). *Science Achievement In The Primary School Years. IEA's Third*

International Mathematics and Science Study. Chestnut Hill, MA: TIMSS International Study Center, Boston College.

Mehrens, W. A. (1992). "Using performance assessment for accountability purposes." *Educational Measurement: Issues and Practice, 11* (1): 3-9, 20.

Meisels, S. J., Dorfman, A., y Steele, D. (1995). "Equity and excellence in group-administered and performance-based assessments." En: M. T. Nettles y A. L. Nettles (Ed.). *Equity And Excellence In Educational Testing And Assessment*. Boston: Kluwer: 243-261.

Mullis, I. V. S., Martin, M. O., Beaton, A. E., Gonzalez, E. J., Kelly, D. L., y Smith, T. A. (1997). *Mathematics Achievement In The Primary School Years. IEA's Third International Mathematics and Science Study*. Chestnut Hill, MA: TIMSS International Study Center, Boston College.

Nwana O. C. (1996). "What are national assessments and why do them?" En: P. Murphy, V. Greaney, M. E. Lockheed, E. C. y Rojas (Ed.). *National Assessments. Testing The System*. Washington, DC: Banco Mundial: 21-31.

OCDE (1995). *Education At A Glance*. París: Organización para la Cooperación y el Desarrollo Económico.

Olivares, J. (1996). "Inclusive national testing: Chile's 'Quality of education assessment system'." En: A. Little y A. Wolf (Ed.). *Assessment In Transition. Learning, Monitoring And Selection In International Perspective*. Oxford: Pergamon: 118-133.

Phillips, G. W., Mullis, V. S., Bourque, M. L., Williams, P. L., Hambleton, R. K., Owen, E. H., y Barton, P. E. (1993). *Interpreting NAEP Scales*. Washington, DC: National Center for Education Statistics.

Rojas, C. (1996). "The Colombian education assessment system." En: P. Murphy, V. Greaney, M. E. Lockheed y C. Rojas (Ed.). *National Assessments. Testing The System*. Washington, DC: Banco Mundial: 147-155.

Ross, K. N. (1987). "Sample design." *International Journal of Educational Research, 11*, 57-75.

UNESCO (1990). *World Declaration on Education for All. Meeting Basic Learning Needs*. Nueva York: Autor.

United States, National Center for Education Statistics. (1991). *Education Counts: An Indicator System To Monitor The Nation's Educational Health*. Washington, DC: Autor.

Weiss, C. H. (1979). "The many meanings of research utilization." *Public Administration Review,* 39: 426-431.

Woodhouse, G., y Goldstein, G. (1996). "The statistical analysis of institution-based data." En: H. Goldstein y T. Lewis (Ed.). *Assessment: Problems, Developments And Statistical Issues*. Chichester, UK: Wiley: 135-144.

CAPÍTULO 3

EVALUACIÓN Y ESTÁNDARES DE CURRÍCULO

Gilbert A. Valverde

El capítulo anterior trató el tema de la evaluación de los resultados del aprendizaje escolar a escala nacional. El capítulo siguiente hace un puente entre las aspiraciones nacionales o metas educativas y los resultados obtenidos, por medio del seguimiento y la evaluación de los procesos de la educación. Estos procesos se centran en torno del currículo escolar, es decir, en lo que se enseña y en cómo se enseña. El autor discute las características y ventajas de un modelo sistémico e ilustra la discusión con base en la información y las metodologías desarrolladas a propósito del Tercer Estudio Internacional de Matemáticas y Ciencias, TIMSS. El capítulo concluye con un análisis de las implicaciones de esta perspectiva para la política educativa en campos tales como los textos escolares, el currrículo, los indicadores de desarrollo de la educación y las estrategias de reforma.

INTRODUCCIÓN

Se está llevando a cabo en las Américas una reforma arrolladora de la educación. Cierto número de países han emprendido distintas políticas encaminadas a reformar a fondo los sistemas educativos existentes (Álvarez, 1997). Cabe advertir, en particular, el aspecto *sistémico* de muchos movimientos de reforma educativa en la región, ya que tiene repercusiones profundas para la formulación de la política educativa en general y, en especial, para el diseño de los sistemas de evaluación y seguimiento. El movimiento contemporáneo de reforma educativa defiende el punto de vista de que las reformas necesarias requieren políticas integrales dirigidas a los múltiples y diversos aspectos del sistema educativo. Ciertamente, los programas dedicados a promover la reforma educativa en América Latina han defendido este punto de vista (véase Slavin, 1994), y está documentado en distintas descripciones de reformas que actualmente se realizan en toda la región (por ejemplo, en Álvarez y Ruiz-Casares, 1997). Las políticas sistémicas—es decir, las destinadas a reorganizar la autoridad y los recursos entre los individuos y las entidades a fin de modificar el sistema por el que se imparte la educación—presentan un reto formidable tanto a los responsables

de la política como a los evaluadores.

Es vital para el futuro de la política de reforma educativa comprender qué características estructurales de un sistema educativo son esenciales para las actividades de reforma y los efectos de estas mismas. La comprensión de cómo la política educativa puede modificar estas estructuras y cuáles serán los resultados probables, mejora enormemente al considerar las metas, procesos y los productos educativos como un sistema integrado. En pocas palabras, para comprender mejor cómo cambiar o reformar la educación, los responsables de la política han de tener a su disposición los indicadores para medir la educación *sistémicamente*.

Una medición significativa de los sistemas educativos requiere un marco conceptual integral y una serie correspondiente de indicadores concebidos para relacionar las distintas partes del sistema entre sí y sus resultados. Si se desea que las tecnologías para la evaluación de la reforma progresen, necesitamos metodologías para recuperar datos rigurosamente, de forma uniforme y sobre un número significativo de factores asociados con los *procesos* de la escolaridad—no meramente sobre metas y productos—. Esto hace posible el análisis significativo de las opciones normativas, lo que resulta en una comprensión más profunda de los sistemas existentes así como de las alternativas posibles. Está claro que, en esta encrucijada, es vital para muchos países latinoamericanos considerar los sistemas de indicadores a la luz de su potencial para esclarecer la política. La política actual que se ha dictaminado o se está considerando en la región incluye normas para el establecimiento de metas desde una entidad centralizada (la mayoría de las veces, un ministerio o una autoridad provincial/regional análoga) o estrategias formuladas desde los niveles inferiores concebidas para promover la innovación pedagógica en la escuela o en el aula. Con frecuencia, ambos tipos de normas coexisten en reconocimiento de la importancia que tiene la innovación pedagógica para revitalizar los sistemas educativos en un contexto en el que las autoridades nacionales o regionales tienen aún que asegurar normas mínimas en esos sistemas educativos.

Independientemente de si un país aplica las normas dictadas desde los niveles superiores o formuladas desde los niveles inferiores—o una combinación de ambas clases—, los sistemas de indicadores desempeñan un papel importante. No obstante, su importancia está directamente relacionada con el hecho de si proporcionan o no información que permita a los responsables de la política evaluar los méritos relativos de las distintas políticas. En la actualidad, la mayor parte de la formulación de la política educativa tiene por fin aumentar, entre otras cosas, la eficiencia y calidad de la educación. Dicho contexto requiere indicadores que no sólo evalúen lo que aprenden los alumnos sino cómo se les enseña y los tipos de resultados obtenidos. Así pues, los responsables de la política no sólo requieren indicadores de los resultados obtenidos por los alumnos sino también indicadores de los procesos que expliquen los resultados a fin de evaluar las opciones que puedan coadyuvar al logro de sus metas.

En el presente documento se examinará inicialmente cómo el seguimiento de las

políticas relacionadas con las metas educativas—tales como los estándares de currículo—y los productos educativos (mediciones del rendimiento de los alumnos) conducen inevitablemente a una evaluación que considera la importancia de los *procesos* que transforman las metas en productos. Luego, se presentan las razones para contemplar nuevos modelos de evaluación que hagan de la relación entre estas dos dimensiones normativas el centro explícito del proceso de evaluación. Entonces, los procesos educativos se convierten en un vínculo integral entre las metas y los resultados que han de tenerse presentes en los sistemas de indicadores.

Después se abordarán algunas de las características esenciales de un modelo sistémico. Éste es el modelo sistémico de las oportunidades educativas resultantes del trabajo del Estudio de Oportunidad en Matemáticas y Ciencias (SMSO) en la Universidad Estatal de Michigan para el Tercer Estudio Internacional de Matemáticas y Ciencias (TIMSS) concluido en fecha reciente por la Asociación Internacional para la Evaluación de los Resultados Educativos (IEA). Estas características tienen por fin orientar la elaboración de evaluaciones *integradas* del currículo, metas y prácticas de instrucción y rendimiento de los alumnos que constituyen una innovación en el campo de la evaluación educativa.

El documento delineará la forma en que el estudio de los componentes curriculares puede proporcionar a los responsables de la política indicadores valiosos en un clima de reforma educativa. También explicará cómo se organizan los sistemas educativos para proporcionar el currículo a los alumnos. Este documento explorará estas innovaciones metodológicas, detallará sus aportes a la evaluación de la política educativa y proporcionará ejemplos sacados de los datos recopilados para estudios realizados por el SMSO y el Centro Nacional de Investigación de los Estados Unidos para el TIMSS.

POLÍTICAS EDUCATIVAS, SISTEMAS DE INDICADORES Y EL PROCESO INVISIBLE

Desde principios de la década de 1980, los países de las Américas han emprendido la formulación de políticas encaminadas a subsanar las deficiencias percibidas de los actuales sistemas educativos. Dos resultados importantes de estas iniciativas normativas y las correspondientes estrategias de evaluación han sido el desarrollo de sistemas de evaluación y la reformulación de las políticas existentes de dirección curricular o la formulación de nuevas políticas en esta esfera. Las políticas relacionadas con la evaluación y las metas son concebidas a menudo como mecanismos conjuntos de control del sistema o "gestión del sistema" (Apple, 1990) en el que se mantiene que la medición de los productos (la mayoría de las veces mediante pruebas de rendimiento) contra fines (estándares de currículo) proporciona una evaluación científicamente válida y políticamente suficiente.

Sin embargo, se ha recopilado un cúmulo cada vez mayor de pruebas convincentes en la investigación sobre la evaluación y la gerencia del currículo en relación con los problemas, tanto en las pruebas como en la política curricular, que tienen

repercusiones importantes para la formulación de la política orientada a la reforma.

Escollos en la evaluación y la reforma basada en el contenido

Por más de tres décadas, muchos países de todo el mundo han realizado una serie de evaluaciones internacionales y nacionales del rendimiento educativo. Todos han declarado, en mayor o menor grado, que estas evaluaciones han proporcionado datos importantes útiles para determinar la eficacia de los sistemas educativos. La política educativa en América Latina es un reflejo de estas tendencias internacionales. Esto es evidente, por ejemplo, en una descripción reciente de la evaluación en una propuesta para el programa curricular de Colombia: *La evaluación busca detectar fallas y aciertos para incluir los correctivos necesarios que garanticen el progreso del alumno* (Ministerio de Educación Nacional de Colombia, 1990).

En muchos sistemas contemporáneos, un papel primordial (si no único) de la evaluación en el examen de la eficacia de los programas y políticas es el uso de las pruebas del rendimiento de los alumnos. La hipótesis implícita de dichos enfoques para la evaluación es la de que las puntuaciones sobre el rendimiento de los alumnos pueden atribuirse en realidad a las experiencias educativas de los alumnos y, por tanto, representan una evaluación válida de la eficacia comparativa de los programas y las políticas educativos.

Cada vez más, se ha argumentado que las evaluaciones, en particular las nacionales, pueden ser los motores principales que impulsen la reforma educativa. Ciertamente, esto ocurre en debates recientes en Estados Unidos, incluido el llamamiento del Presidente Clinton para un programa nacional de pruebas. También se observa claramente en la expansión de las iniciativas latinoamericanas en materia de pruebas. El esquema deductivo implícito que conduce a dichas políticas parece tener los siguientes componentes:

- Pruebas clave que esclarecerán y especificarán explícitamente las metas de la educación.
- Pruebas para medir si los maestros, las escuelas y los alumnos alcanzan con éxito estas metas.
- Al hacer responsables a los maestros, las escuelas y los alumnos del uso de estas evaluaciones, se les proporcionará la motivación necesaria para mejorar la enseñanza y el aprendizaje.

Se ha declarado que un esquema deductivo como éste se fundamenta en la psicología y pedagogía del comportamiento (Noble y Smith, 1994) y, por tanto, concuerda con las iniciativas de reforma más actuales en las Américas. Muchas reformas promulgan el uso de enfoques pedagógicos modernos encaminados a enseñar importantes aptitudes de razonamiento y pensamiento críticos, algunas recomendando incluso el uso de pedagogías constructivistas (Ministerio de

Educación Nacional de Colombia, 1990; Ministerio de Educación Pública de Costa Rica, 1996; Secretaría de Educación Pública de México, 1993). Los programas de evaluación *de sistema cerrado*, al concentrarse en la equiparación de las metas con los resultados, ignoran el *proceso* educativo. Suponen implícitamente que el proceso de escolaridad "se desarrollará" una vez que se aclaren las metas en las pruebas. En realidad, estos regímenes de evaluación ignoran paradójicamente la *escolaridad*, es decir, el proceso mediante el cual se alcanzan las metas de la instrucción en el aula. Hacen responsables a los maestros, los alumnos y las escuelas de los resultados, ignorando el hecho de si éstos tienen o no control sobre los factores que dan a los niños la oportunidad de aprender las aptitudes evaluadas.

Un análisis reciente de muchos estudios de pruebas pone de manifiesto que dichas políticas afectan a los alumnos y a los maestros de formas incompatibles con los movimientos de reforma actuales. Documenta que dichas políticas de evaluación han creado entornos en los que los maestros ignoran temas que las pruebas dejan sin abarcar, descuidan enfoques de enseñanza en equipo, subrayan aptitudes básicas frente a aptitudes mentales más elevadas—e incluso equiparan el formato de su enseñanza con el formato de las pruebas (Noble y Smith, 1994).

Por tanto, nos vemos confrontados con la cuestión vital de si es posible que dichas evaluaciones identifiquen las fallas y los puntos fuertes *relacionados con las experiencias educativas* de los niños que permitirían identificar posibles medidas correctivas.

La esencia de este desafío radica en el interrogante de si estas evaluaciones miden o no, en realidad, lo que declaran medir (Airasian y Madaus, 1983). El problema que afrontamos es el del justo equilibrio de las evaluaciones en cuanto a la *instrucción* y el *currículo*. Dicho simplemente, el problema está en determinar si las pruebas miden los objetivos del currículo y si las escuelas proporcionan a los alumnos instrucción en las aptitudes y el conocimiento evaluados. La respuesta a este problema representa un elemento vital para determinar la pertinencia normativa de las evaluaciones y constituye la esencia misma del empleo de las evaluaciones en la política educativa.

Los estudios de la IEA introdujeron la noción de la Oportunidad para Aprender (OTL) como medio de asegurar la validez técnica de sus conclusiones (McDonnell, 1995) y, al hacerlo, introdujeron no sólo un concepto que revolucionó la tecnología de la evaluación educativa sino que, también, proporcionó una herramienta para realzar la pertinencia normativa de las evaluaciones.

La OTL proporcionó a los estudios de la IEA una medida del grado de correspondencia existente entre las pruebas administradas y las prácticas educativas en cada uno de los países participantes. Principalmente, se pidió a los maestros que mirasen una copia de las preguntas en las evaluaciones y dijesen si habían proporcionado o no a sus alumnos la instrucción necesaria para contestar esas preguntas correctamente. La simplicidad de estas medidas iniciales no expresa la

enorme importancia conceptual de la OTL pero fue un primer paso en reconocer la necesidad de caracterizar la *instrucción* a fin de explicar los resultados obtenidos. La OTL fue la medida de las oportunidades necesarias para desempeñarse bien en las pruebas de rendimiento y, por tanto, un reconocimiento de que las prácticas de instrucción de los maestros contribuyen al rendimiento de los alumnos.

En esencia, la OTL es la medida del proceso. Las evaluaciones tradicionales eran incapaces de medir la serie de experiencias de instrucción que explican el conocimiento y las aptitudes que medían. Los instrumentos de la OTL comenzaron como medidas del contenido abarcado por los maestros de los alumnos que participaron en las evaluaciones. Incluso en los sistemas educativos más centralmente dirigidos, hay una amplia variación en los temas y aptitudes abarcados por los maestros. Los investigadores han demostrado que la oportunidad de aprender las aptitudes sometidas a prueba es una variable explicativa significativa del desempeño de los alumnos (Burnstein, 1993; Burstein y colaboradores, 1990; McDonnell, 1995; Muthén y colaboradores, 1995). Desde la aparición inicial de la OTL como medida del contenido abarcado en el aula, este concepto, y las consiguientes tecnologías de medición, se ha hecho más complejo y más capaz de presentar la variación en las oportunidades educativas proporcionadas a los niños (Guiton y Oakes, 1995; McDonnell, 1995; Schmidt y colaboradores, 1996; Schmidt y colaboradores, 1997a; Shavelson y Webb, 1995).

La OTL proporciona el vínculo vital entre metas y resultados que convierten las meras pruebas en una evaluación educativa más sólida. *Reconoce que la forma en que se utilizan los recursos educativos es ciertamente tan importante como los recursos utilizados*. La inclusión de medidas de la OTL en las evaluaciones educativas proporciona indicadores del proceso educativo a medida que se desarrolla en las escuelas y aulas y facilita a los responsables de la política las herramientas vitales para la evaluación y formulación de políticas de reforma. Esta preocupación con las metas y los procesos es sumamente importante dado el carácter especial de las actuales políticas de reforma educativa.

Una serie importante de iniciativas de investigación y política encaminadas a mejorar la educación en las Américas la constituye el movimiento de reforma escolar sistémica impulsada por el contenido. Una cantidad considerable de trabajo en los Estados Unidos ha contribuido a apoyar este tipo de reforma (véase Clune, 1993; O'Day y Smith, 1993). El principio central de estas iniciativas es el hecho de que la reforma de los sistemas educativos ha de ser impulsada por el *currículo*. Se declara que han de formularse currículos ambiciosos y que han de diseñarse mecanismos apropiados para la aplicación de estos currículos, de forma que los alumnos puedan tener la oportunidad de alcanzar altos niveles de rendimiento. La reforma basada en el contenido mantiene que un currículo básico proporciona una base para determinar qué recursos se necesitan para garantizar que se proporciona a los alumnos las oportunidades requeridas para dominar el contenido. Así pues, el currículo surtiría efecto directamente en la formación y certificación de los maestros, los cursos ofrecidos por las escuelas, los recursos de instrucción y las

estructuras de dirección y responsabilidad educativas. En realidad, a veces se mantiene que *se requiere* contrastar las prácticas curriculares y educativas en el aula para seguir con exactitud cómo se proporciona a los alumnos una oportunidad adecuada de aprender, y es una de las responsabilidades de evaluación de un sistema reformado.

No hay duda de que en los países latinoamericanos hay altas expectativas en relación con el papel de las normas curriculares; ilustran este hecho las declaraciones del actual Presidente de México mientras que prestaba servicios como Secretario de Educación: "El plan y los programas de estudio son un medio para mejorar la calidad de la educación..." (Zedillo Ponce de León, 1993).

Los teóricos de la reforma sistémica basada en el contenido hacen referencia a sistemas educativos con fuertes currículos nacionales al formular sus propuestas normativas. Por ejemplo, O´Day y Smith (1993) declaran, para el caso de los Estados Unidos, que: "Cuando se implante plenamente, este modelo de reforma sistémica basada en el contenido sería una adaptación singularmente americana de las políticas y estructuras educativas de muchos de los países altamente desarrollados del mundo" (página 252). Clune (1993) también indica que un resultado posible de la reforma escolar sistémica basada en el contenido sería un sistema centralizado de dirección y currículos "que se asemeja a los currículos nacionales y los ministerios de educación de países que tienen altos niveles de rendimiento de los alumnos" (página 233). Los problemas surgen de cierto número de casos en los que los estándares se formulan solamente como metas pedagógicas del sistema, sin consideración a la pedagogía en sí que conduce a su realización. Cuando dichas normas se combinan en programas de evaluación de *sistema cerrado*, como indicamos anteriormente, las consecuencias para la reforma pueden ser desastrosas.

Como hemos visto, la OTL surgió como respuesta a una cuestión muy fundamental relacionada con la justicia (que para los técnicos de evaluación se denomina *validez* de las evaluaciones). ¿Es justo enjuiciar a los alumnos contra el desempeño en pruebas para las cuales no les han preparado sus experiencias educativas? ¿Es justo enjuiciar a los maestros o a las escuelas contra el desempeño de sus alumnos en tales evaluaciones cuando no tienen control sobre los recursos de instrucción necesarios para impartir las aptitudes requeridas? También existe una situación política educativa análoga a la OTL. Ésta la constituyen las normas relacionadas con la provisión de oportunidades para aprender.

A veces conocidas como *estándares de enseñanza escolar* (McDonnell, 1995; Porter, 1993), tienen la posibilidad de realzar la eficacia de la política curricular desempeñando dos funciones muy importantes (Porter, 1993), una simbólica, otra técnica.

La función simbólica de las normas de prestaciones escolares puede ser proporcionar una nueva imagen o paradigma, un sentido de "posibilidad

imaginativa" (Apple, 1992) en relación con lo que pudiera ser la instrucción ideal. Bien formuladas, dichas imágenes pueden motivar a los maestros a ser innovadores y creativos. La función técnica guarda relación con el seguimiento de la implantación. Las normas de enseñanza escolar también pueden proporcionar la base para reglamentar las acciones de los maestros, los funcionarios escolares y otras entidades educativas y políticas.

Entre las estrategias contemporáneas de reforma figura un hincapié en enseñar conocimiento real o algorítmico menos simplista y aptitudes más conceptuales y críticas. Es evidente que dicho hincapié requiere, en el caso de muchos sistemas educativos, cambios no sólo en *lo que* se enseña sino también en *cómo* se enseña. Naturalmente, esto constituye un aliciente primordial para que los responsables de la política consideren tanto el marco de las metas curriculares como el establecimiento de normas de enseñanza.

Un principio fundamental de los defensores de la política educativa sistémica es el de que ha de ser impulsada por el currículo. Se mantiene que han de formularse currículos ambiciosos y que han de diseñarse mecanismos apropiados para implantar dichos currículos, de forma que los alumnos tengan la oportunidad de alcanzar altos niveles de rendimiento. La reforma basada en el contenido sostiene que un currículo básico proporciona los medios de establecer normas de enseñanza escolar. Así pues, el currículo influiría directamente en la formación y certificación de los maestros, los cursos escolares ofrecidos y los recursos y las estructuras de instrucción del gobierno y responsabilidad educativa (incluidos el diseño y la administración de pruebas). En realidad, se mantiene que es necesario contrastar el currículo con las prácticas educativas en el aula para supervisar con exactitud la forma en que se proporciona a los alumnos una oportunidad adecuada para aprender y que constituye una responsabilidad de evaluación de un sistema reformado.

Así pues, las estrategias contemporáneas de reforma, así como las tecnologías de evaluación, requieren examinar mucho más de cerca los procesos de educación de forma que aumenten la utilidad de la información al servicio de las estrategias de reforma. El dilema, tal como se presenta en la Figura 1, consiste en formular modelos de evaluación que abran la "caja negra" de los procesos educativos a fin de comprender cómo se transforman en productos los insumos y las metas.

PROCESO VISIBLE: UN MODELO SISTÉMICO DE OPORTUNIDADES EDUCATIVAS

En una era de reforma educativa, los responsables de la política de dicha reforma requieren evaluación pertinente para la política que sea capaz de descubrir las correlaciones institucionales y pedagógicas de las diferencias en los niveles de resultados. Tanto los responsables de la política como los investigadores están llegando a la conclusión de que el papel adecuado de la evaluación educativa consiste en ayudar a descubrir características sistémicas, prácticas de instrucción y

Figura 1. El problema de integrar el proceso en las estrategias de evaluación

Nota: El reto para la evaluación es concebir los procesos de tal forma que los vinculen directamente con las metas, los insumos y los productos.

otros insumos y procesos educativos que mejor se presten a configurar los resultados educativos deseados.

Al diseñar las estrategias de recolección y análisis de datos para el TIMSS, se utilizó un estudio de los sistemas educativos tal como existen actualmente para impartir experiencias educativas a los alumnos. Se construyó una serie compleja de colecciones interrelacionadas de datos mediante un modelo sistémico de oportunidades educativas, en el cual los datos sobre resultados son sólo un componente importante. Concebimos estas oportunidades educativas como la formulación de intenciones nacionales o subnacionales y su aplicación en el aula. Son instrumentos de política y técnicas docentes concebidas para guiar las experiencias realizadas. El modelo reconoce que el aprendizaje propiamente dicho o las experiencias educativas realizadas son procesos subjetivos asimilados por cada alumno. El punto de partida para la formulación del modelo lo constituye un modelo tripartito del currículo, tal como se ilustra en la Figura 2, que ha sido la característica tradicional de los estudios de la IEA por algún tiempo. Este modelo hace una distinción analítica entre el currículo como metas del sistema, como

Figura 2: Modelo analítico del currículo de la IEA

instrucción y como rendimiento del alumno. Cada una de estas dimensiones se conoce, respectivamente, como el currículo intencional, el implantado y el logrado.

El modelo sistémico de oportunidad educativa, formulado por el Centro TIMSS de los Estados Unidos en virtud de una donación separada conocida como Estudio de Oportunidad en Matemáticas y Ciencias (SMSO)[1], gira en torno a la premisa de que un fin de un sistema educativo, y centro primordial de su papel pedagógico, consiste en permitir a los educadores y responsables de la política proporcionar oportunidades para aprender las ciencias y las matemáticas. En este esquema, reconocemos los currículos, las escuelas y los maestros como elementos que, a la vez, definen y delimitan las posibles experiencias de aprendizaje proporcionadas a los alumnos para aprender matemáticas y ciencias. Las oportunidades para alcanzar una meta de aprendizaje (por ejemplo, tener conocimiento útil y aptitudes centradas en temas concretos) surten un efecto directo sobre el logro real de la meta por parte de los alumnos.

El término *meta* se utiliza para referirse al contenido, tal como lo definen marcos curriculares analíticos (Robitaille y colaboradores, 1993; Survey of Mathematics and Science Opportunities, 1992a; Survey of Mathematics and Science Opportunities, 1992b) e incluye expectativas de desempeño y actitudes. Esta sección presenta las características más destacadas del modelo sistémico SMSO de oportunidades educativas que es el marco teórico para la recolección de datos del proyecto TIMSS (Schmidt y colaboradores, 1996; Schmidt y McKnight, 1995). Este modelo está organizado en torno a cuatro preguntas fundamentales: (1) ¿Qué se espera que aprendan los alumnos? (2) ¿Quién imparte la instrucción? (3) ¿Cómo se organiza la instrucción? y (4) ¿Qué han aprendido los alumnos?

Figura 3. Modelo sistémico de oportunidad educativa del SMSO

Nota: Este modelo recupera el proceso educativo mediante vínculos explícitos con insumos, metas y productos. Estos vínculos se establecen mediante el sistema de categorías de los marcos curriculares analíticos del SMSO.

La Figura 3 ilustra la construcción y las relaciones recíprocas en el modelo. Las cuatro columnas representan las cuatro preguntas principales de la investigación y las cuatro hileras especifican los cuatro niveles del sistema educativo que se examinarán. Las flechas muestran la red de relaciones entre los distintos elementos.

¿Qué se espera que aprendan los alumnos?

La descripción de la oportunidad educativa comienza con la descripción de qué conocimiento y aptitudes se supone que dominen los alumnos. Hay tres niveles principales del sistema educativo en el que comúnmente se establecen las metas: el nivel nacional o regional, el de la escuela y el del aula. La primera cuestión entraña determinar no sólo las metas de aprendizaje para un sistema o país en su conjunto sino la diferenciación de dichas metas para subdivisiones de los sistemas educativos, entre ellas, las regiones, la separación de los alumnos por especialización o rendimiento, diferentes tipos de escuelas y las aulas. Concentrándose en cada nivel del sistema educativo, las preguntas generales se convierten en:

1. ¿Qué metas se trata de alcanzar al nivel nacional o subnacional?
   a. ¿Cómo se alinean estas metas con los grados y cuál es su secuencia?
   b. ¿Quién decide qué metas se alcanzarán y cómo se adoptan estas decisiones?
   c. ¿Cómo y sobre qué base se diferencia a los alumnos con respecto a las metas que han de tratar de alcanzar?
2. ¿Qué metas se trata de alcanzar al nivel de escuela (o para los distintos tipos de escuela) y cómo influyen en ellas las metas nacionales o subnacionales?
3. ¿Qué metas se trata de alcanzar al nivel de aula o para los distintos tipos de aula de los alumnos? ¿Qué influencia ejercen los objetivos del currículo a nivel nacional, subnacional o escolar?

Las cuestiones sustantivas aquí giran en torno al currículo. Siguiendo el lenguaje utilizado por la IEA, las metas de aprendizaje especificadas al nivel nacional/ subnacional las constituye el currículo intencional, mientras que la especificación de las metas de aprendizaje al nivel de escuela o aula lo constituye, al menos en parte, lo que se denomina el currículo implantado. El currículo como metas y planes para distribución de experiencias potenciales de aprendizaje, es uno de los conceptos principales de este modelo. Tal como sugieren las preguntas anteriores, el TIMSS estudia el concepto en cada nivel del sistema educativo.

¿Quién imparte la instrucción?

La oportunidad educativa también se caracteriza por la provisión de instrucción y por las personas a cargo de impartir dicha instrucción. Éste es el papel central de los maestros. Una forma de caracterizar el papel de los maestros consiste en describir las cualificaciones para la certificación oficial de los maestros. Entre los aspectos de esta certificación figuran la limitación en cuanto a grados y materias, los requisitos de resultados educativos para el otorgamiento de una licencia, el tipo de grado educativo requerido y cualquier trabajo específico del curso o experiencia práctica

que pueda exigirse adicionalmente.

El entorno profesional en el que trabajan los maestros y la organización social que les rodea, también influyen en la provisión de la instrucción. Esto guarda relación con aspectos tales como la proporción del tiempo profesional empleado en enseñanza propiamente dicha y la proporción del tiempo empleado durante la jornada escolar en planificación del trabajo del curso.

Puesto que los maestros desempeñan un papel tan central, es vital especificarlos minuciosamente. Muchos estudios han mostrado la forma en que las características de los maestros influyen en la calidad de la instrucción y, por tanto, en la calidad de la oportunidad educativa (Carlsen, 1991; Coyen, 1988). Dos amplias categorías de características de los maestros son especialmente importantes: los antecedentes y las ideas del maestro. Entre las variables relativas a los antecedentes del maestro figuran la edad, el género, la formación, la materia enseñada y la experiencia docente. Las ideas del maestro guardan relación con la orientación a la materia y la pedagogía específica de la materia. Las ideas de los maestros en torno a la materia pueden afectar a las prácticas de instrucción y a rendimiento de los alumnos. Estas ideas relacionadas con la materia se refieren al punto de vista que tiene un maestro sobre las disciplinas que enseña. Las ideas pedagógicas, por el contrario, se refieren a las ideas de los maestros acerca de la mejor forma de enseñar un tema en particular dentro de una disciplina.

La investigación para esta sección se basa en un examen de las características personales de los maestros: sus ideas acerca de las materias escolares y cómo se enseñan óptimamente y cómo se capacitan. Todas estas características influyen en la transmisión de los currículos o de las experiencias de aprendizaje potenciales. Esto nos permite examinar los tipos de alineación que existen entre la formación del maestro y la práctica, y los objetivos del currículo nacional/subnacional, proporcionando así datos que, cuando se combinan con otros sobre resultados, pueden someter a prueba la hipótesis de los defensores de la reforma sistémica en el sentido de que dichas alineaciones inciden en los resultados educativos. También ayudará identificar determinadas ideas pedagógicas y sobre la materia que asegurarán óptimamente la transmisión de currículos interesantes a los alumnos a través de todo el sistema.

¿Cómo se organiza la instrucción?

Junto con las expectativas para aprendizaje de los alumnos y la transmisión de la instrucción por los maestros, la forma en que se organiza la instrucción influye en el currículo implantado y, por tanto, en las experiencias potenciales de aprendizaje de los alumnos. Ningún responsable de la política en particular determina la organización de la instrucción. La serie de posibles lugares de toma de decisiones en relación con la instrucción es amplia. Las decisiones pueden adoptarse en la cúspide de la jerarquía educativa o en los niveles intermedios. El lugar de la toma de decisiones también puede ser la escuela—con los administradores de escuela o los

maestros en el aula a cargo de decidir cómo enseñan el contenido a sus alumnos—.

Esta difusión de la autoridad decisoria se manifiesta de muchas formas. Por ejemplo, hay variantes importantes en la estructura de edad-grados de los sistemas educativos y en la naturaleza de las escuelas que sirven a series diferentes de grados. Además, el sistema escolar organiza a menudo a los alumnos en categorías curriculares diferentes. No hay duda de que estas circunstancias determinan las características organizativas del sistema educativo. Influyen también en las cualificaciones de la fuerza docente y en los tipos de recursos de instrucción de que disponen esos maestros. Además, tienen una gran influencia en el tiempo y los recursos materiales a disposición de la población estudiantil.

Los cursos escolares ofrecidos están implícitos en la forma en que se organiza la instrucción. Las funciones que desempeñan las escuelas en apoyo de los cursos ofrecidos en distintas materias escolares y los papeles que los individuos desempeñan en apoyo de esas funciones surten efecto en la organización de la instrucción.

La organización de la instrucción también ocurre en aulas individuales. Así pues, entre los factores que influyen en el currículo implantado figuran elementos tales como: el uso de textos, la estructura de las lecciones, los usos del material de instrucción, la evaluación de los alumnos, su participación, las tareas y la agrupación en clase de estos últimos.

La organización de las metas nacionales y subnacionales en programas de estudio y en los cursos ofrecidos son decisiones que se adoptan a menudo a nivel de escuela. Por tanto, las autoridades escolares son fuentes potenciales importantes de información sobre las oportunidades docentes. Los maestros pueden proporcionar datos extensos sobre sus prácticas de instrucción y sobre cómo organizan y enseñan las lecciones a los alumnos.

¿Qué han aprendido los alumnos?

Siendo el fin de la definición del sistema educativo identificar los factores que determinan qué aprenden los alumnos y cómo lo aprenden, es importante examinar las características de los alumnos. Amén de la influencia de las metas del currículo, los maestros y la organización de la instrucción, las características de los alumnos influyen en el proceso mediante el cual se ponen en práctica las posibles experiencias de aprendizaje. Las escuelas no pueden proporcionar experiencias reales de aprendizaje, sólo oportunidades para dichas experiencias. En última instancia, el interés, las motivaciones, las aspiraciones y otras características de los alumnos, influenciados por sus antecedentes personales, transforman las experiencias potenciales en experiencias realizadas. Entre las características importantes de los alumnos figuran las siguientes: el historial académico del alumno; la situación económica de su familia; y su condición socioeconómica; la participación cultural de la familia; el concepto de sí mismo; el tiempo empleado

fuera de la escuela; la motivación y el interés.

No es posible ni deseable identificar y medir cada uno de los componentes de un sistema educativo. Sin embargo, este modelo de experiencias educativas reconoce las conexiones recíprocas entre los elementos del sistema educativo de forma análoga a las conceptualizaciones de quienes proponen el movimiento de reforma sistémica en los Estados Unidos (Clune, 1993; O´Day y Smith, 1993). Este es un modelo genérico que puede utilizarse para describir muchos sistemas educativos concretos. No defiende un sistema en particular sino que constituye, más bien, un patrón contra el cual se identifican diferentes variantes sistémicas. Su finalidad principal es el estudio de los sistemas educativos.

UNA ESTRATEGIA DE EVALUACIÓN SISTÉMICA

Después de presentar el modelo de experiencias educativas potenciales, analizaré ahora una estrategia de medición, elaborada también por el SMSO, concebida para vincular el modelo con distintas colecciones de datos. Esta estrategia, elaborada para el estudio entre países de la educación en matemáticas y ciencias, fue concebida intencionalmente para aplicarla a la evaluación nacional, ya que el grupo de investigación estuvo convencido desde el principio de que las *evaluaciones entre países* sólo son útiles si también son *evaluaciones nacionales* útiles.

La subdivisión del currículo en tres componentes (véase la Figura 1) tiene por fin servir de instrumento analítico, una forma de aislar los aspectos del currículo para estudio. Sin embargo, las colecciones de datos del TIMSS están concebidas de forma integrada. Los elementos integrantes son los marcos curriculares del TIMSS. El TIMSS vincula directamente el currículo y los elementos de la materia del modelo a través de los marcos curriculares. Los marcos proporcionan un idioma unificador para el TIMSS, en su conjunto y para el componente de análisis curricular, en particular. Estos marcos representan una especificación de categorías y aspectos múltiples del contenido y expectativas y perspectivas sobre desempeño en una forma pertinente para la medición del material curricular y las pruebas. Proporcionan una serie de convenciones que sirven de sistema de lenguaje común. En pocas palabras, representan la operación de las "matemáticas y las ciencias" como entidades medibles en la provisión de oportunidades de aprendizaje.

Los marcos curriculares del TIMSS se utilizan para caracterizar todos los componentes de las colecciones de datos del TIMSS. Los cursos, las preguntas de pruebas, las escalas de resultados, los textos, las lecciones de las guías del currículo y las ideas pedagógicas de los maestros se ilustran todos en términos de categorías de marcos y pueden vincularse a través de los análisis. Por ejemplo, podemos vincular la cobertura del currículo en guías de este último y los textos sobre el tema "Estimar computaciones" (una categoría sacada del aspecto del contenido de los marcos de matemáticas del TIMSS) con el desempeño de los alumnos en relación con una serie de preguntas de rendimiento que abarcan este tema. También vemos la forma en que las expectativas sobre desempeño que se requieren para resolver la

serie concreta de preguntas, que pudieran ser, por ejemplo, "seleccionar y construir objetos matemáticamente equivalentes" (una categoría sacada del aspecto de expectativas de desempeño del marco), están vinculadas con la presencia de dichas expectativas de desempeño en los textos o guías curriculares. También podemos vincular la respuesta de los maestros a las ideas pedagógicas acerca de enseñar la "proporcionalidad" con lecciones reales sobre proporcionalidad y con una presentación mediante textos de ese contenido.

El uso de estos tipos de marcos analíticos para vincular entre sí cada instrumento de la colección de datos proporciona una oportunidad singular para estudiar la alineación de todos los elementos del sistema al facilitar oportunidades para aprender elementos curriculares definidos. Estos marcos constituyen la tecnología central concebida para vincular medidas de intención, implantación y resultados.

La medición de la intención

El modelo sistémico de oportunidad educativa del SMSO reconoce la importancia de las especificaciones nacionales o subnacionales de las metas de aprendizaje, que se entienden incorporan el currículo intencional. Dichas exposiciones de las metas de aprendizaje de los alumnos se hallan a menudo en documentos oficiales tales como las guías curriculares y los programas de estudio. Sin embargo, estos documentos varían en el nivel de detalle con que especifican las metas de aprendizaje. Aunque no siempre son oficiales, los textos también proporcionan información sobre las metas de aprendizaje de los alumnos que es habitualmente detallada y combinan información acerca de las metas al nivel oficial o semioficial con información sobre la probabilidad de distintas metas al nivel de escuela y de maestro.

El currículo, al especificar las metas de aprendizaje al nivel nacional o subnacional, establece parámetros que subrayan ciertas experiencias potenciales de aprendizaje y limitan otras. Por ejemplo, en un país con un currículo oficial obligatorio promovido por evaluaciones nacionales y una dirección de inspección de escuelas, la inclusión de una meta de aprendizaje no garantiza su cobertura—es decir, que la oportunidad se proporcionará realmente en las aulas—, pero sí aumenta considerablemente la probabilidad de que eso ocurra. De forma análoga, la ausencia de una meta aumenta la probabilidad de que no se proporcionen experiencias de aprendizaje potencial relacionadas con dicha meta, pero, al igual que antes, tratamos sólo de cambios en probabilidades—en las distribuciones de probabilidad de experiencias potenciales de aprendizaje—y no de la certidumbre de que se proporcione la oportunidad.

Las diferencias entre las provincias, los municipios o las escuelas en la especificación de las metas de aprendizaje, y las políticas relacionadas con las metas de aprendizaje, tienen una importancia primordial para comprender estas relaciones. La especificación a nivel de sistema de las metas de aprendizaje establece parámetros mediante los cuales se limitan experiencias de aprendizaje potenciales, aunque

quizás no en grado igual, independientemente del tipo de sistema.
El modelo contempla el currículo como elemento de definición principal de las experiencias educativas potenciales. Configura las metas y expectativas nacionales de aprendizaje. Estas experiencias potenciales son en sí características sistémicas importantes, y están estrechamente vinculadas con la política educativa y las cuestiones de la calidad de la educación. La caracterización de estas cualidades proporciona una comprensión de los tipos de filosofías y metas y constituye un elemento fundamental para comprender el contexto en el que se implantan los currículos.

En consecuencia, es un resultado de importancia equivalente al de otro indicador: el rendimiento de los alumnos. Cómo se diseña el currículo pretendido, qué incluye, qué políticas de control guardan relación con él y quién decide lo que se espera que aprendan los alumnos son características fundamentales de política de los sistemas educativos. Estas características son una preocupación central en toda formulación de políticas de reforma. La capacidad de vincular estas características con otros resultados como las prácticas de instrucción y el rendimiento de los alumnos es otro de los puntos fuertes de este modelo: nos permite estudiar la forma en que las experiencias educativas potenciales se transforman y llevan a la práctica a medida que se implanta el currículo en las escuelas y aulas.

Entre los objetivos del TIMSS figura una serie de preguntas sobre las relaciones de los conceptos ilustrados en la Figura 1, que trascienden del rendimiento. Las metas nacionales/subnacionales constituyen uno de estos conceptos. Puesto que establecen los parámetros o restricciones dentro de los cuales se proporcionan experiencias potenciales de aprendizaje, una caracterización de estas restricciones constituye una parte central como contexto para interpretar la provisión de oportunidades educativas. Los responsables de la política, los encargados de formular el currículo y los educadores en muchos países luchan con la reforma de sus currículos. ¿Qué debería tener un "currículo de nivel internacional", por ejemplo, en matemáticas y ciencias como contenido y como expectativas de desempeño? ¿Debería incluir metas no cognoscitivas también? ¿Qué consiguientes políticas necesitan estar asociadas con dicho currículo? Si los responsables de la política desean seriamente abordar estas cuestiones a medida que emprenden la reforma, entonces se necesitan descripciones integrales y basadas en documentos de lo que los diferentes países están haciendo realmente para iniciar dichas deliberaciones.

El primer objetivo del análisis del currículo del SMSO consiste en responder, a nivel del sistema, a la pregunta: ¿Qué se espera que aprendan los alumnos en matemáticas y ciencias? Tal como se describió anteriormente, la especificación de la intención o las metas nacionales/subnacionales delinea las características que definen la oportunidad educativa a nivel más amplio. La inclusión o exclusión de un contenido o desempeño concreto en las metas nacionales/subnacionales incide en las probabilidades asociadas con la provisión de dichas experiencias potenciales de aprendizaje. De esta forma, el currículo intencional proporciona la condición

necesaria—pero sólo en un sentido estocástico—para la provisión de oportunidad educativa. El estudio de dichos parámetros facilita la base fundamental a partir de la cual es interpretable una estructura de oportunidades de un país.

Sirven de ejemplo los vínculos importantes del currículo intencional y el currículo implantado. El examen de estos vínculos es equivalente a la exploración de la estructura de oportunidad educativa. El aspecto de estos vínculos, el papel que desempeñan las guías curriculares y textos y la forma en que las políticas, tales como los sistemas nacionales de control, la organización a nivel de grado y la separación de los alumnos por especialización y rendimiento afectan a los vínculos, son todas cuestiones importantes en el estudio de experiencias educativas potenciales.

Además, desde el punto de vista de los ministerios nacionales o las autoridades subnacionales, se necesita una caracterización desde una perspectiva internacional de la intención nacional/subnacional, junto con las consiguientes políticas como base para las decisiones normativas. Cuáles son las metas nacionales/subnacionales y qué políticas de implantación están relacionadas con ellas son cuestiones inherentes en el establecimiento de la política a nivel nacional o subnacional. En muchos casos, la mayoría de los responsables de la política han dejado ya de querer fijar dichas políticas de forma aislada, ignorando lo que otros, en particular sus competidores económicos, están haciendo. Desde estos dos puntos de vista—un estudio de la oportunidad educativa y el establecimiento de la política curricular a nivel de ministerio o subnacional—, la caracterización del currículo como respuesta a la pregunta de qué se espera que aprendan los alumnos es esencial. Sin embargo, también lo son las consiguientes preguntas de quién determina las metas de aprendizaje en un país, en qué forma y con qué estructuras documentales se especifican las metas, qué políticas existen en apoyo de las metas de aprendizaje y en qué forma están interrelacionados estos factores del sistema educativo.

La pregunta tomada del modelo sistémico de oportunidades educativas del SMSO de que se ocupa éste, el primer aspecto del análisis curricular, es la siguiente: ¿Qué se espera que aprendan los alumnos en matemáticas y ciencias a nivel nacional o subnacional?

La medición de la implantación

Las escuelas y maestros, mediante sus características y actividades, también ayudan a configurar las experiencias potenciales de aprendizaje proporcionadas a los alumnos. La organización curricular del sistema y la escuela, y las características y el conocimiento de la materia por parte de los maestros, también inciden en la provisión y calidad de las experiencias potenciales de aprendizaje. Las metas de aprendizaje utilizadas por el maestro en el aula y las materias del curso que proporciona la escuela delimitan y configuran adicionalmente estas oportunidades. Así pues, el paso siguiente en la comprensión de la oportunidad educativa consiste en medir qué ocurre en la provisión de la instrucción a los alumnos. Es decir,

¿cómo se implanta el currículo intencional en las escuelas y aulas? *¿Cuáles son las medidas concretas adoptadas por las personas u organizaciones para garantizar el cumplimiento real de la política curricular?*

La implantación del currículo entraña toda la gama de operaciones realizadas por los maestros y los administradores de escuela—las personas a las que se les encomienda la implantación de la política nacional—. Así pues, las metas escolares, las metas de contenido del maestro, la oportunidad de aprender y las prácticas de instrucción son todos aspectos importantes de implantación. Otras características importantes son la fuerza docente, las escuelas y el sistema nacional/subnacional que ejercen una gran influencia en la calidad de la implantación.

El SMSO ha diseñado instrumentos para medir las metas de contenido del maestro, las metas de la escuela y medidas más clásicas de la OTL. Para las metas de contenido del maestro, el instrumento mide qué pretenden enseñar los maestros a los alumnos y cuántos períodos se asignan a la instrucción del contenido. Los métodos miden las metas de contenido contra los marcos curriculares del TIMSS y los vinculan con otros aspectos de la implantación y con las medidas del currículo intencional y el currículo logrado.

El aspecto de OTL de la instrumentación ha sido rediseñado sustancialmente con respecto a estudios anteriores de la IEA y se ha vinculado a través de los marcos curriculares a todas las demás medidas. La labor de rediseño entraña identificar grupos de preguntas de las pruebas del rendimiento como ejemplos de categorías de contenido diferentes a partir de los marcos curriculares del TIMSS. Se pide a los maestros que respondan si han enseñado o pretenden enseñar el contenido representado por las preguntas y presentado en términos de los marcos. Este nuevo diseño tiene por fin aumentar la probabilidad de que los maestros proporcionen datos de OTL que no se concentren en los aspectos específicos de preguntas individuales sino, más bien, en las categorías de contenido que éstos representan.

La implantación tiene también aspectos cualitativos. Las características personales de quienes tienen a cargo la responsabilidad de cumplir las intenciones nacionales en el aula determinan la implantación. Las características de las escuelas y el sistema educativo en el que tiene lugar la implantación también influyen en ésta. Distintas medidas evalúan los criterios que han de cumplir los maestros potenciales para ser habilitados oficialmente para desempeñarse como educadores y ejercer la profesión en un marco en particular (tipos de escuelas, trabajo con ciertos tipos de alumnos, etc.). Otras partes del instrumento miden el entorno profesional que los maestros experimentan en sus escuelas y exploran distintas características personales de los maestros. Estas características incluyen en especial sus orientaciones a la materia e ideas pedagógicas relacionadas con la enseñanza de las materias escolares.

La provisión de actividades de instrucción y de experiencias potenciales de aprendizaje ocurre principalmente en la escuela local y, más exactamente, en el aula bajo la guía de un maestro en particular. Debido al carácter "práctico" de estas

actividades y a la capacidad limitada de la mayoría de los sistemas para seguir las decisiones de las escuelas o los maestros individuales, existen entre las escuelas y entre las aulas dentro de las escuelas, incluso en los sistemas "centralizados", importantes variantes en la instrucción. En consecuencia, los estudios de las experiencias potenciales de aprendizaje han de concentrarse principalmente en las prácticas de instrucción—las acciones propiamente dichas realizadas en el aula—.

En contraste, puesto que las metas son intenciones, no experiencias proporcionadas, pueden establecerse en cualquier nivel, o en todos los niveles de un sistema. Así pues, un sistema "centralizado" pudiera significar un sistema en el que se especifican las metas nacionales y en el que se hacen grandes esfuerzos por controlar las actividades a nivel de aula, o pudiera significar un sistema en el que existen tales metas pero se permite autonomía en muchos niveles al impartir la instrucción. Es sumamente importante al nivel tanto conceptual como empírico distinguir entre las metas de las actividades y las experiencias reales. Por tanto, al examinar cómo se organiza la instrucción, el instrumento[2] del TIMSS explora las siguientes series de preguntas:

1. ¿Cómo está organizado el sistema al nivel nacional o subnacional?
    a. ¿Qué niveles de grados tiene el sistema?
    b. ¿Cómo se agrupan los grados en las escuelas (primaria, secundaria inferior, secundaria superior, etc.)?
2. ¿Qué papeles y funciones desempeñan las escuelas en la provisión de la instrucción y cómo se organizan las matemáticas y las ciencias en el programa escolar?
    a. ¿Cuáles son los programas de estudio?
    b. ¿Se enseñan las materias en secuencia o paralelamente?
    c. ¿Qué papeles desempeñan las escuelas en apoyo de la instrucción (por ejemplo, asignación de recursos frente a relaciones con la comunidad, los alumnos y la familia?
    d. ¿Cómo se asignan los recursos a la instrucción?
    e. ¿Cómo guarda esto relación con las metas de aprendizaje de los maestros?
3. ¿Cómo se organiza la instrucción en el aula?
    a. ¿Cuál es la estructura de una lección?
    b. ¿Cómo se utilizan en la instrucción los textos y guías curriculares?
    c. ¿Qué actividades de instrucción se emplean, tales como laboratorios, conferencias, demostraciones, agrupaciones, evaluaciones de alumnos, etc.?
    d. ¿Cómo se utilizan en la instrucción la práctica en clase y las tareas?
    e. ¿De qué formas participan los alumnos en la instrucción?
    f. ¿Cuál es la relación con las metas de aprendizaje de los maestros?

La organización de la instrucción ocurre tanto en la escuela como en las aulas individuales. El propio sistema tiene una estructura mediante la cual se organiza la instrucción en un sentido más amplio. Esto último tiene profundas repercusiones para la organización de la instrucción a nivel de escuela y de aula y, en conjunto, desempeña un papel importante en determinar la calidad de las experiencias

potenciales de aprendizaje.

Medición del rendimiento

En el modelo sistémico de oportunidades educativas, la medición del rendimiento entraña medir no sólo el dominio que han adquirido los alumnos sino también otras características de estos últimos. Al igual que es necesario distinguir entre intención (o experiencia potencial) e implantación (actividades relacionadas con la definición adicional de experiencias potenciales), también es necesario distinguir entre experiencias potenciales y la participación activa de un alumno en una experiencia y el rendimiento (los cambios producidos en el alumno, tal como se reflejan en pruebas, mediante la actividad y como resultado de la actividad de aprendizaje realizada). Por tanto, las características de los alumnos, entre las que figuran sus antecedentes sociales, culturales y económicos, inciden en el rendimiento ya que afectan a lo que experimenta un alumno en particular durante una actividad. Estos factores de los antecedentes del alumno que median entre las "experiencias potenciales" y el rendimiento del alumno se consideran, por razones de seguridad, parte de la estrategia de medición del rendimiento o el currículo logrado, ya que tienen una fuerte relación con el rendimiento obtenido por los alumnos en la mayoría de los países.
Para el TIMSS, el SMSO utilizó datos curriculares preliminares para redactar las especificaciones de contenido para estas pruebas a fin de asegurar una concentración en las áreas de contenido que son más importantes para los países participantes.

Existen tres clases de preguntas en estas pruebas: las de opciones múltiples; las que requieren respuestas breves, tales como un número, una frase, un diagrama o una oración; y las de respuesta extensa que requieren un esfuerzo más prolongado para demostrar por escrito el razonamiento del alumno, los pasos intermedios, las explicaciones escritas u otros productos más extensos del desempeño del alumno. Además, hay distintas tareas de desempeño que se administrarán a una subserie de la muestra nacional de alumnos.

Una estrategia de análisis

El examen de los sistemas educativos entraña la descripción de los elementos más importantes como metas y la forma en que se ejecutan en el aula y realizan en el rendimiento del alumno. Así pues, la estrategia de análisis ha de integrar los componentes sistémicos asociados con los currículos intencional, implantado y logrado.

La clave para esta integración en la estrategia de análisis del SMSO la constituyen los marcos curriculares analíticos. Estos marcos son la estrategia de medición central del estudio. Se utilizan para categorizar los materiales del currículo, enfoques pedagógicos específicos para el contenido y desempeños de los alumnos entre otras medidas de cumplimiento del estudio. Estos instrumentos importantes hacen

posible la interrelación detallada de las distintas colecciones de datos. Éste es uno de los aportes más importantes del SMSO a la evaluación sistémica: la posibilidad de relacionar las formulaciones normativas de las intenciones del currículo con la implantación en la escuela (organización de la escuela y procesos del aula) para alcanzar el currículo (el rendimiento del alumno). Los marcos curriculares hacen esto posible y se está formulando una serie integral de procedimientos analíticos nuevos para integrar los resultados de las distintas mediciones. Tal estrategia integradora tiene por fin descubrir la relación del currículo intencional con la implantación por el maestro y los resultados de los alumnos. El Cuadro 1 presenta un ejemplo de un cuadro de análisis concebido para ilustrar los datos que pueden utilizarse para estudiar dichas relaciones entre los países (o subsistemas).

Este cuadro esboza los tipos de datos que pueden obtenerse en el TIMSS mediante las mediciones descritas en las tres secciones precedentes, por medio de los cuales pueden explorarse las relaciones entre los currículos intencional, implantado y logrado. Ilustra las interesantes posibilidades del diseño innovador del TIMSS. Podemos observar patrones de intenciones nacionales en cuanto se relacionan con la presentación por los maestros del contenido y el rendimiento de los alumnos. La relación de estos patrones con otras características y políticas sistémicas tales como la separación de los alumnos por especialización o por grupos de rendimiento, por ejemplo, proporciona la posibilidad de comprender el cambio sistémico.

La exploración y descripción de cómo los indicadores del currículo intencional guardan relación con las del currículo implantado y el currículo logrado entraña la formulación y el perfeccionamiento de nuevos índices y procedimientos estadísticos conexos. Esta serie de indicadores y sus vínculos con los resultados coadyuvan a alcanzar la meta de formular un sistema sólido de indicadores que reconozca la plena complejidad del proceso educativo.

Este es el aporte central del SMSO a los estudios nacionales de la política educativa—un intento integral por descubrir características sistémicas, sus relaciones recíprocas entre sí y con el rendimiento de los alumnos—un modelo enriquecido para evaluación educativa. La oportunidad de examinar un sistema de indicadores interrelacionados con relaciones explícitas permite examinar precisamente las relaciones que han estado ausentes en muchos de los métodos precedentes de evaluación.

ALGUNOS EJEMPLOS ADICIONALES DE LA APLICACIÓN DE LAS TÉCNICAS DEL SMSO

Una mirada más detenida a la medición del currículo intencional

Las fuentes de información utilizadas por el TIMSS para medir el currículo intencional son las guías curriculares, los textos y los expertos en el currículo. Las descripciones completas de estos métodos pueden hallarse en una serie de libros recientes (Schmidt y Knight, 1995; Schmidt, McKnight, Valverde, Houang y Wiley,

1997b; Schmidt, Raizen, Britton, Bianchi y Wolfe, 1997c).

Cuadro 1. Cuadro de análisis de oportunidad para aprender

Nota: Este es un ejemplo que ilustra algunas de las clases de vínculos que pueden establecerse entre el análisis de los datos del TIMSS, utilizando contenido matemático o de ciencias como elemento vinculador. Presenta algunas de las posibilidades que existen para explorar las relaciones entre la política curricular, las prácticas de instrucción de los maestros y el rendimiento de los alumnos. Tales datos existen para la totalidad de los casi 50 países que componen el TIMSS. (Este cuadro da cuenta sólo de los datos de ensayos de campo para matemáticas de octavo grado en los Estados Unidos.) El cuadro indica que hay algunas áreas de contenido en las que un número mayor de niños resuelven las preguntas de pruebas sobre resultados correctamente cuando sus maestros han abarcado la materia y ésta ha recibido también una atención considerable en los textos y en las guías curriculares del estado. También indica algunas áreas de contenido en las que, a pesar de la cobertura por el maestro y una atención considerable en los textos y las guías curriculares, los alumnos muestran un desempeño deficiente en preguntas de pruebas afines. Utilizando cuadros como éste, junto con la información detallada en las prácticas de instrucción del maestro y el diseño de textos, los evaluadores y los responsables de la política pueden descubrir elementos conducentes a la formulación de una política curricular más eficaz.

A fin de comprender los ejemplos siguientes, es importante advertir que el SMSO diseñó una serie de procedimientos de análisis curricular que incluyó un análisis minucioso de las guías curriculares y los textos. Los procedimientos también

incluyen seguir la cobertura de los temas a través de todos los niveles de grados y cuestionarios para los expertos en el currículo. Estos procedimientos están vinculados con el sistema de lenguaje de los marcos curriculares del TIMSS. Para el procedimiento conocido como análisis documental, se utiliza una metodología de análisis del contenido que entraña dividir los documentos en bloques relativamente homogéneos, cuya sustancia se codifica luego de acuerdo con los marcos. Esto resulta en un inventario detallado del contenido, en expectativas de desempeño y perspectivas a través de todo el libro. Esto se hace para las guías curriculares y los textos en las tres poblaciones. Para un limitado número de temas (llamados temas a fondo), codificadores en el país analizan los documentos para proporcionar información acerca de la cobertura de los temas en todos los niveles de grados, desde el comienzo de la escolaridad hasta el final de la escuela secundaria. Esta tarea también se realiza para todos los temas en los marcos, tomando como base el enjuiciamiento más que un análisis formal de los documentos. Se pidió a expertos (utilizando documentos, según procediese) que proporcionasen esta información. Al Cuestionario de Expertos, que aborda cuestiones más amplias tales como las reformas, los patrones de dirección, la distribución de autoridad en el proceso de toma de decisiones curriculares y el uso de la calculadora y computadora, respondieron a una serie de expertos en matemáticas y ciencias de cada país participante en el TIMSS. Esta recopilación de datos ha concluido y 48 países recuperaron y presentaron datos de más de 1.200 textos y guías curriculares al SMSO.

El caso del currículo, la enseñanza y el rendimiento en Colombia[3]

La República de Colombia participó en un TIMSS que acaba de concluir y los resultados sobre rendimiento, en términos de su comparación con los otros 40 países participantes en este aspecto del estudio, no pueden haber sido más desalentadores. En ciencias, por ejemplo, de 41 países, 39 tenían puntuaciones medias de rendimientos que eran notablemente más altas que las de Colombia. Sólo un país, Sudáfrica, ocupó una clasificación más baja. La situación para el rendimiento en matemáticas en Colombia fue idéntica (Beaton y colaboradores, 1996a; Beaton y colaboradores, 1996b).

Si el único producto del TIMSS fuesen estas puntuaciones de rendimiento, la participación de Colombia habría sido en gran medida un desperdicio. Sin embargo, debido a que el TIMSS utilizó una estrategia de medición basada en un modelo sistémico que incorporaba datos sobre metas, procesos y resultados, hay una serie de análisis pertinentes para la política que los responsables de la política e investigadores colombianos pueden efectuar dentro de los datos.

Para este ejemplo, examinemos primero algunos elementos del currículo intencional de Colombia, tal como se manifiestan en los textos de los alumnos. La Figura 4 presenta datos sobre los énfasis temáticos relativos en una muestra nacionalmente representativa de textos de ciencias de octavo grado en Colombia, así como en otros tres países latinoamericanos del TIMSS: la Argentina (provincia de

Figura 4: Temas de ciencias más subrayados en los textos del octavo grado de los cuatro países latinoamericanos

Nota: Las barras ilustran el porcentaje de espacio de texto dedicado a hasta cinco de los temas más subrayados. Estos datos se derivan de una muestra nacionalmente representativa de todos los textos requeridos o más comúnmente utilizados para el currículo completo obligatorio de ciencias en este nivel de grados.

Figura 5: Temas más subrayados por los maestros

Nota: Estos son los cinco temas más subrayados por la muestra nacional de maestros de octavo grado en Colombia, el Japón y Alemania. A pesar de la presencia de un material considerable sobre ciencias físicas en sus textos (véase la Figura 4), los maestros colombianos parecen dedicar poco tiempo de instrucción a estos temas en relación con las ciencias vitales.

Chaco), la República Dominicana y México.

Como puede verse, los textos colombianos subrayan las ciencias tanto naturales como físicas, situación que concuerda con las metas nacionales enunciadas en la política curricular nacional.

¿Qué ocurre con los procesos del aula? ¿Qué enseñan los maestros en Colombia? Tal como ilustra la Figura 5, una muestra nacional de maestros de octavo grado de Colombia declaran que subrayan las ciencias naturales en su instrucción alineada de cerca con los textos que emplean. Sin embargo, no subrayan las ciencias físicas de manera que concuerde con los textos. No hay duda de que estas variantes en la oportunidad para aprender tienen consecuencias en el rendimiento de los alumnos. El Cuadro 2 ilustra claramente este aspecto.

En este cuadro puede verse claramente que los alumnos en Colombia tienen un desempeño mucho mejor en aquellas subáreas de las ciencias subrayadas en los textos y en la instrucción en el aula. Se requerirá sin duda un análisis adicional en Colombia para establecer las prioridades de la investigación para factores que expliquen la renuencia de los maestros a poner en práctica los aspectos de las ciencias físicas del currículo. Es probable que una comparación de los datos de la enseñanza, los textos y las guías curriculares indique alternativas en la política curricular que tienen posibilidad de garantizar que los alumnos reciben oportunidades apropiadas para aprender las ciencias físicas.

Cuadro 2. Clasificación de resultados de Colombia en el TIMSS de acuerdo con los temas de ciencias

Nota: Los temas en los recuadros son aquellos en los que se hace hincapié en los textos y en las prácticas de instrucción de los maestros, tal como se advirtió en ilustraciones previas.

El caso de los exámenes para alumnos salientes en el Oriente Medio y el Norte de África

Desde la formulación de las metodologías originales para el TIMSS, el SMSO ha tenido ocasión de aplicarlas a nuevos marcos de evaluación, uno de los cuales fue un análisis de los exámenes de conclusión de la escuela en biología y matemáticas en algunos países del Oriente Medio y el Norte de África (Valverde y Schmidt, 1995).

Los exámenes, tal como se indicó anteriormente en este documento, tienen por fin a menudo impulsar la enseñanza y el aprendizaje en los sistemas educativos. Esto se cumple especialmente en el caso de evaluaciones de "entrada", tales como las que se efectúan para proporcionar a los alumnos credenciales al salir de la escuela o determinar su paso a instituciones de educación superior. De estos tipos de prueba, se cumple en particular lo siguiente:

> ... las pruebas ilustran las metas educativas... nuestras pruebas, en especial en las etapas posteriores del proceso educativo, han de reflejar y formar a la sociedad en general, *promoviendo y definiendo sus criterios de liderazgo, así como articulando nuestras aspiraciones para ella*. (Wiley, 1982: 100) (las cursivas son mías)

El Banco Mundial, al preparar un informe sobre educación en el Oriente Medio y África del Norte (MENA), deseaba investigar lo que los exámenes al salir de la escuela, evaluados utilizando métodos de SMSO, podían revelar acerca de la calidad de la educación. Hacia tal fin, se recopiló una muestra de pruebas.

Los análisis realizados por el SMSO se concentraron en el estudio y la comparación de las siguientes características de los exámenes: (1) la materia de matemáticas o biología abarcada en cada pregunta de prueba y (2) las expectativas asociadas para desempeños de los alumnos o las aptitudes completas que se pretendía que los alumnos mostrasen al responder con éxito a cada pregunta.

Cada uno de estos formularios de prueba se codificó utilizando métodos analíticos de contenido formulados por el SMSO para el componente de análisis curricular del TIMSS. La unidad de análisis que se codificó fue cada *tarea de prueba*. El término se utiliza en vez de *pregunta de prueba* ya que muchas preguntas de prueba individualmente numeradas en cada prueba están constituidas por series de tareas discretas. Es decir, las preguntas incluyen a veces más de una pregunta a la que el examinando tiene que responder—cada una de estas preguntas se denominó una tarea de prueba—.

Cada tarea en cada formulario para la que se destinó una respuesta del alumno separada se codificó utilizando categorías procedentes de los marcos curriculares del TIMSS. Los códigos asignados a cada tarea o respuesta pretendida (a menudo más de uno) constituyen los datos de los que se crearon archivos de análisis. Se emplearon

dos categorías de códigos: *contenido* y *expectativas de desempeño*. *Contenido* se refiere simplemente a las áreas del tema de matemáticas o ciencias abarcadas por cada pregunta o tarea. *Expectativa de desempeño* se refiere a los desempeños previstos de los alumnos, desempeños que se espera que los alumnos demuestren cuando participan en la solución de las tareas de prueba. Esta última categoría de códigos especifica qué pretendían las pruebas que los alumnos hiciesen con el contenido de matemáticas o ciencias. Se realizaron distintos análisis. Aquí, quisiera destacar un análisis que tiene importantes repercusiones para la política educativa en MENA y que ilustra la pertinencia normativa de los indicadores formulados por el SMSO que no están estrictamente relacionados con los métodos tradicionales de evaluación.

Se identificaron para cada prueba las expectativas de desempeño, es decir, lo que las pruebas pretendían que los alumnos hiciesen con la biología o las matemáticas. Los perfiles creados se perfeccionaron luego para identificar expectativas predominantes (las que representan el mayor énfasis de las pruebas). La definición operativa de predominio fueron las expectativas de desempeño presentes al menos en 70 por ciento de las tareas de prueba. El Cuadro 3 presenta un resumen que compara exámenes de matemáticas a la salida de la escuela con los de un país europeo que tiene una larga tradición en el uso de tales pruebas: el examen de bachiller de Francia administrado en París y en Aix en 1991 y 1992[4].

Cuadro 3. Expectativas de desempeño básico

Notas: Este cuadro presenta expectativas de desempeño básico en los exámenes de salida de la escuela en matemáticas de la región de MENA contra exámenes análogos en Francia. Se reflejan metas muy diferentes en términos de las competencias requeridas.

La lista de expectativas sobre puntuaciones de MENA incluye cuatro expectativas de las áreas generales de *conocer, utilizar procedimientos de rutina y razonamiento matemático*, tal como se caracterizan en los marcos del TIMSS. La mayoría de éstas también se hallan presentes en las pruebas francesas sacadas de la muestra. Los formularios franceses también incluyeron expectativas de desempeño en las áreas de *investigación y resolución de problemas y comunicación* que no fueron comunes en la muestra de MENA considerada en su conjunto.

Éste y otros análisis realizados en este estudio condujeron al descubrimiento de una característica sorprendente de las pruebas MENA. Al examinar el contenido y las expectativas de desempeño, el uno con las otras, hallamos una ausencia relativa de tareas que conecten las matemáticas y los contextos del mundo real. Esto es extraordinario ya que ésta es un área que ha recibido una atención considerable en las iniciativas de reforma educativa de matemáticas en Europa, América del Norte y la propia MENA. En realidad, las tareas que evalúan las capacidades de los examinandos en el área de resolución de problemas estuvieron ausentes, en gran medida, en las pruebas de la muestra de MENA. Estas pruebas parecen indicar un concepto de matemáticas escolares como una materia dedicada principalmente al reconocimiento y la repetición de definiciones y teoremas y la realización de algoritmos y otros procedimientos de rutina. Ciertamente, hubo diferencias entre los formularios de prueba y los países respecto a este énfasis, pero la tendencia general parece clara.

La política curricular enunciada en estos países establece claras prioridades para la promoción de la capacidad de pensamiento crítico, la resolución de problemas prácticos y las aptitudes de proceso de la información, como ocurre en el Reino de Jordania (Billeh, 1996). Sin embargo, a pesar de su elevada importancia, los exámenes a la salida de la escuela no dan prioridad a estas aptitudes, haciendo que sea sumamente improbable que reciban alta prioridad cuando los maestros y los alumnos sean juzgados con respecto a ella.

El descubrimiento de estos tipos de faltas de concordancia entre dos aspectos del currículo intencional—pruebas y política curricular—ha resultado ser útil para cada país en particular y para el Banco Mundial a medida que evalúan las opciones normativas para reformar la educación de modo que propicie el desarrollo de economías competitivas.

ALGUNAS CONCLUSIONES: EVALUACIÓN Y FORMULACIÓN DE UNA POLÍTICA DEL CURRÍCULO EFICAZ

Varias hipótesis causales implícitas sirven de base al interés público y político en la educación en todas las Américas: mantienen que los sistemas mejorados de enseñanza llevarán a un mayor rendimiento por parte de un número más elevado de alumnos. Esto, a su vez, se considera como el resultado de un número mayor de alumnos mejor preparados que entran a la fuerza laboral, lo que resultará en un desempeño económico más competitivo en la economía global. Se entiende que un

conocimiento general mayor de los temas escolares estimulará las innovaciones a través de todos los sectores de la sociedad y la economía y resultará en un electorado y una población de consumidores más informados, en mayor productividad de la fuerza laboral en general y en una competitividad mejorada en la industria nacional. Este criterio ha sido expuesto enfáticamente y repetidas veces—la educación es una de las inversiones humanas más importantes que puede hacer un país—(Álvarez, 1997; Becker, 1964; Board of International Comparative Studies in Education, 1993; Ministerio de Educación Nacional de Colombia, 1990; Ministerio de Educación Pública de Costa Rica, 1996; el Banco Mundial, 1995).

Amén de este argumento que vincula el desarrollo educativo personal a través de la escolaridad con el crecimiento y la competitividad económica nacional, el debate sobre la política educativa nos ha llevado a considerar las instituciones educativas como un sistema.

Este sistema se considera susceptible de reforma mediante la formulación de una política informada. Políticos en diferentes países—así como organizaciones internacionales, tales como la Organización de Cooperación y Desarrollo Económicos y la UNESCO—han establecido una serie de prioridades de evaluación.

Los enfoques para la evaluación educativa que incluyen el proceso educativo en su consideración de las metas y resultados educativos pueden efectuar un aporte importante a la formulación de la la política educativa. Esto es posible al proporcionar indicadores que nos permitan comprender los vínculos entre las metas nacionales, lo que ocurre en las escuelas y aulas y los resultados de los alumnos. La importancia normativa de un enfoque, tal como el SMSO, radica en que se concentra en la noción de oportunidad educativa como resultado importante de los sistemas educativos y el que ofrezca un modelo de oportunidad educativa que permita su descripción en términos sistémicos.

En las páginas precedentes, he descrito e ilustrado algunos de los beneficios de la estrategia de evaluación sensible al currículo formulada por el SMSO. También he explorado la forma en que los indicadores sobre proceso pueden proporcionar información pertinente para la formulación de la política. Algunos de los beneficios más importantes que ofrece el sistema de indicadores aquí descrito guarda relación con la capacidad creciente de formular una política curricular eficaz.

La política curricular está sometida en la actualidad a revisiones sustantivas en muchos países. Uno de los problemas más importantes que deben enfrentar los responsables de la política quizás sea el de reconciliar los mensajes diversos relativos al currículo enviados por sus instrumentos más importantes: normas o marcos curriculares, programas de estudio, textos y evaluaciones del rendimiento.

El sistema de medidas descrito aquí permite al evaluador caracterizar, con un nivel considerable de detalle, la forma en que cada uno de estos instrumentos normativos

transmite mensajes relacionados con el currículo a cada persona y entidad en el sistema. Puede revelar, por ejemplo, si los objetivos establecidos en los marcos curriculares están respaldados por pruebas de rendimiento y textos, proporcionando un instrumento importante para los sistemas en los que algunos elementos de la política curricular son íntegramente la responsabilidad de las autoridades centrales (tales como los marcos curriculares) y otros elementos se dejan en manos de otras personas (por ejemplo, los textos por publicarse en editoriales privadas).

Proporcionan, por tanto, un instrumento ya sea para formulación de políticas coherentes relativas a la elaboración de textos (para los sistemas en los que las autoridades centrales elaboran los textos), adopción (para los sistemas que producen una lista de textos "autorizados" comercialmente elaborados) o proporcionan términos de referencia útiles en situaciones en que los países compran textos dentro de un programa de licitación competitiva.

Los marcos curriculares, si bien establecen metas educativas importantes para el sistema, no están concebidos en general para sugerir estrategias de ejecución específicas. Dichas estrategias de ejecución (es decir, sugerencias de acciones concretas que van a emprender organizaciones o personas específicas a fin de alcanzar metas pedagógicas) están contenidas en otros instrumentos normativos tales como los programas de estudio y textos. El diseño de textos y programas de estudio más eficaces también lo realzan sistemas de evaluación tales como el concebido por el SMSO. Puesto que la evaluación de los procesos educativos proporciona información sobre prácticas de instrucción y sobre cómo éstas afectan al logro de los distintos objetivos pedagógicos, constituyen una buena base para el diseño de dichos instrumentos. La inclusión de tales medidas en los programas nacionales de evaluación proporciona un mecanismo de reacción vital que será útil para los diseñadores del currículo, redactores de textos, elaboradores de pruebas y otras personas en el sistema, en la modificación y perfeccionamiento de programas de estudio, textos y pruebas durante un período de tiempo.

Obviamente, estos indicadores también son útiles en la concepción de políticas eficaces relativas a la formación de maestros antes y durante el ejercicio de su profesión. Un rasgo extraordinario de la política curricular actual en gran parte del continente americano es que, a pesar de un llamamiento a un cambio considerable en la pedagogía, hay poca consideración explícita de cómo el sistema educativo va a producir los maestros requeridos para impartir los currículos más exigentes. Al revelar estrategias de ejecución eficaces e ineficaces—en cuanto corresponden a metas pedagógicas específicas relacionadas con el rendimiento esperado de los alumnos y el contenido de las materias escolares—tales indicadores pueden proporcionar información importante que puede influir en los currículos de preparación de maestros y el diseño de programas de capacitación en el desempeño de las funciones encaminados a corregir las deficiencias en la provisión de oportunidades educativas.

Los sistemas tradicionales de evaluación, al concentrar la atención de los

responsables de la política y el público en las "carreras de caballos" u "olimpiadas de rendimiento" que comparan las calificaciones de las pruebas de los niños, o las puntuaciones medias de las clases, escuelas, provincias—e inclusive de países—materializan dichas categorías y confunden las cuestiones importantes. La conciencia de un rendimiento deficiente puede concentrar la atención en el mejoramiento. Sin embargo, dichas categorías en sí son inútiles, confusas y pueden ser en realidad destructivamente alarmistas. Pueden conducir, en el mejor de los casos, a soluciones propuestas simplistas y, en el peor de los casos, a políticas mal concebidas con consecuencias perniciosas. Lo que requieren los responsables de la política son datos sobre la forma en que los elementos del sistema educativo persiguen una serie de metas educativas, y sobre qué características sistémicas, prácticas de instrucción y otros procesos educativos son los mejores para alcanzar estas distintas metas. Este tipo de información hará una contribución positiva al debate público al ayudarnos a orientar nuestra atención hacia la consideración de políticas con posibilidades de impulsar nuestras metas.

Esto es evaluación tal como se concibe en el moderno análisis de la política: "ampliar la tarea de la evaluación más allá de la mera medición de los resultados frente a sus causas" (Pressman y Wildavsky, 1984: página xv). Vistos de esta forma, tales enfoques de la evaluación educativa pueden proporcionar una fuente rica de información destinada a esclarecer temas clave de política nacional de los años noventa y más allá en todo el continente americano.

NOTAS

[1] El SMSO fue una donación otorgada a la Universidad Estatal de Michigan en su calidad de Centro de Investigación de los Estados Unidos para el TIMSS, que coordinó la investigación y efectuó una labor de desarrollo con los centros nacionales de investigación de un subgrupo de países participantes en el TIMSS: el Japón, Noruega, Suiza, España, Francia y los Estados Unidos. El SMSO tuvo por fin realizar lo siguiente para el TIMSS: una serie de marcos curriculares, un modelo conceptual para orientar la elaboración de instrumentos y el análisis de datos, cuestionarios para maestros y alumnos, técnicas para medir y estudiar los currículos intencionales y las especificaciones (anteproyectos) para las pruebas de rendimiento del TIMSS.

[2] Nuestro libro (Schmidt y colaboradores, 1996) y una serie de informes técnicos del Centro de SMSO de la Universidad Estatal de Michigan presentan informes detallados sobre la instrumentación. Para mayor información, diríjase al autor a la siguiente dirección: SMSO, 457 Erickson Hall, East Lansing, Michigan 48824-1034.

[3] Deseo expresar reconocimiento a mi colega, el Dr. Carlos Jairo Díaz, por autorizar la utilización de los datos del TIMSS de Colombia para estos análisis. Todas las preguntas relacionadas con las circunstancias específicas de la participación colombiana en el TIMSS deben dirigirse al doctor Díaz a la siguiente dirección: Multitaller de Materiales Didácticos, Universidad del Valle, Ciudad Universitaria Meléndez, Apartado Aéreo 25360, Cali, Colombia.

[4] Los datos para estos exámenes franceses los proporcionó el Centro Nacional de Mejoramiento de la Educación Científica, Washington, D.C. Los datos franceses fueron codificados por el Dr. John Dossey, Universidad Estatal de Illinois. Se puede obtener información adicional acerca de estos exámenes en: Britton, Edward D. y S.A. Raizen. *Examining the Examinations for College-Bound Students*. Boston / Dordrecht / Londres, Kluwer Academic Press, 1996.

REFERENCIAS

Airasian, P. W., y Madaus, G. F. (1983). "Linking Testing and Instruction: Policy Issues." *Journal of Educational Measurement, 20* (2): 103-118.

Álvarez, B. (1997). "Naturaleza y contexto de las reformas educativas de final de siglo." En: B. Álvarez y M. Ruiz-Casares (Eds.), *Senderos de cambio: Génesis y ejecución de las reformas educativas en América Latina y el Caribe* (Vol. 1: 1-22). Washington DC: Advancing Basic Education and Literacy Project.

Álvarez, B., y Ruiz-Casares, M. (Eds.). (1997). *Senderos de cambio: Génesis y ejecución de las reformas educativas en América Latina y el Caribe*. Washington DC: Advancing Basic Education and Literacy Project.

Apple, M. W. (1990). *Ideology and Curriculum*. (Second Edition ed.). Nueva York: Routledge.

Apple, M. W. (1992). "Do the standards go far enough? Power, Policy, and Practice in Mathematics Education." *Journal for Research in Mathematics Education, 23*(5): 412-431.

Beaton, A. E., Martin, M. O., Mullis, I. V. S., Gonzalez, E. J., Smith, T. A., y Kelly, D. L. (1996a). *Mathematics Achievement in the Middle School Years: IEA's International Mathematics and Science Study*. Chestnut Hill, MA: TIMSS International Study Center.

Beaton, A. E., Martin, M. O., Mullis, I. V. S., Gonzalez, E. J., Smith, T. A., y Kelly, D. L. (1996b). *Science Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study*. Chestnut Hill, MA: TIMSS International Study Center.

Becker, G. S. (1964). *Human Capital: A Theoretical and Empirical Analysis*. Nueva York: Columbia University Press.

Billeh, V. Y. (1996). *Educational Reform in Jordan: An Analytical Overview*. Amman, Jordania: National Center for Human Resources Development.

Board on International Comparative Studies in Education. (1993). *A Collaborative Agenda for Improving International Comparative Studies of Education*. Washington, DC: National Academy Press.

Burstein, L. (1993). "Studying Learning, Growth, and Instruction Cross-Nationally: Lessons about why and why not engage in cross-national studies." En: L. Burstein (Ed.), *The IEA Study of Mathematics III: Student Growth and Classroom Processes* . Nueva York: Pergamon Press.

Burstein, L., Aschbacher, P., Chen, Z., y Sen, Q. (1990). *Establishing the Content Validity of Tests Designed To Serve Multiple Purposes: Bridging Secondary-*

Postsecondary Mathematics. . Los Angeles: University of California Los Angeles, Center for Research on Evaluation, Standards, and Student Testing,.

Carlsen, W. S. (1991). "The construction of subject matter knowledge in primary science teaching." En: J. Brophy (Ed.), *Teachers' subject matter knowledge and classroom instruction* (Vol. 2). Greenwich, CT: JAI Press.

Clune, W. H. (1993). "The Best Path to Systemic Educational Policy." *Educational Evaluation and Policy Analysis, 15*(3): 233-254.

Cohen, D. K. (1988). "Teaching Practices: Plus que ça change...." En: P. Jackson (Ed.), *Contributing to Educational Change: Perspectives on Research and Practice* (págs. 27-84). Berkeley, CA: McCutchan.

Elmore, R. F., y Fuhrman, S. H. (1994). "Governing Curriculum: Changing Patterns in Policy, Politics, and Practice." En: R. F. Elmore y S. H. Fuhrman (Eds.), *The Governance of Curriculum* (págs. 1-10). Alexandria, VA: Association for Supervision and Curriculum Development.

Guiton, G., y Oakes, J. (1995). "Opportunity to Learn and Conceptions of Educational Equality." *Educational Evaluation and Policy Analysis, 17*(3): 323-336.

Linn, R. L. (1987). *State-by-State Comparisons of Student Achievement: The Definition of the Content Domain for Assessment* (275). Los Angeles: University of California Los Angeles, Center for Research on Evaluation, Standards, and Student Testing.

McDonnell, L. M. (1995). "Opportunity to Learn as a Research Concept and a Policy Instrument." *Educational Evaluation and Policy Analysis, 17* (3): 305-322.

Ministerio de Educación Nacional de Colombia. (1990). *Matemáticas: Marco General y Propuesta de Programa Curricular - Octavo Grado de Educación Básica*. Bogotá: República de Colombia. Ministerio de Educación Nacional.

Ministerio de Educación Pública de Costa Rica. (1996). *Programa de Estudios: Matemática: Educación Diversificada*. San José, Costa Rica: Ministerio de Educación Pública.

Muthén, B., Huang, L.-C., Jo, B., Khoo, S.-T., Goff, G. N., Novak, J. R., y Shih, J. C. (1995). "Opportunity-to-Learn Effects on Achievement: Analytical Aspects." *Educational Evaluation and Policy Analysis, 17*(3): 371-403.

Noble, A. J., y Smith, M. L. (1994). "Old and New Beliefs About Measurement-Driven Reform: 'Build It and They Will Come.'" *Educational Policy, 8*(2): 111-136.

O'Day, J. A., y Smith, M. S. (1993). "Systemic Educational Reform and Educational Opportunity." En: S. H. Fuhrman (Ed.), *Designing Coherent Educational Poliy* (págs. 250-312). San Francisco: Jossey-Bass.

Porter, A. C. (1993, June-July). "School Delivery Standards." *Educational Researcher, 22*: 24-30.

Porter, A. C., Archbald, D. A., y Tyree, A. K., Jr. (1991). "Reforming the Curriculum: Will Empowerment Policies Replace Control?" En: S. H. Fuhrman y B. Malen (Eds.), *The Politics of Curriculum and Testing* (págs. 11-36). Londres: The Falmer Press.

Pressman, J. L., y Wildavsky, A. (1984). *Implementation*. (Third ed.). Berkeley: University of California Press.

Robitaille, D. F., Schmidt, W. H., Raizen, S., McKnight, C., Britton, E., y Nicol, C.

(1993). *Curriculum Frameworks for Mathematics and Science*. (Vol. 1). Vancouver, Canada: Pacific Educational Press.

Schmidt, W. H., Jorde, D., Barrier, E., Gonzalo, I., Moser, U., Shimizu, K., Sawada, T., Valverde, G. A., McKnight, C., Prawat, R. S., Wiley, D. E., Raizen, S. A., Britton, E. D., y Wolfe, R. G. (1996). *Characterizing Pedagogical Flow: An Investigation of Mathematics and Science Teaching in Six Countries*. Dordrecht, Países Bajos: Kluwer Academic Publishers.

Schmidt, W. H., y McKnight, C. C. (1995). "Surveying Educational Opportunity in Mathematics and Science: An International Perspective." *Educational Evaluation and Policy Analysis, 17*(3): 337-353.

Schmidt, W. H., McKnight, C. C., Raizen, S. A., Jakwerth, P. M., Valverde, G. A., Wolfe, R. G., Britton, E. D., Bianchi, L. J., y Houang, R. T. (1997a). *A Splintered Vision: An Investigation of U.S. Science and Mathematics Education*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Schmidt, W. H., McKnight, C. C., Valverde, G. A., Houang, R. T., y Wiley, D. E. (1997b). *Many Visions, Many Aims: A Cross-National Investigation of Curricular Intentions in School Mathematics*. (Vol. 1). Dordrecht, The Netherlands: Kluwer Academic Pulblishers.

Schmidt, W. H., Raizen, S. A., Britton, E. D., Bianchi, L. J., y Wolfe, R. G. (1997c). *Many Visions, Many Aims: A Cross-National Investigation of Curricular Intentions in School Science*. (Vol. 2). Dordrecht, The Netherlands: Kluwer Academic Pulblishers.

Secretaría de Educación Pública de México. (1993). *Educación Básica Secundaria: Plan y Programa de Estudios*. México D. F.: Secretaría de Educación Pública.

Shavelson, R. J., y Webb, N. M. (1995). On Getting It Right. *Educational Evaluation and Policy Analysis, 17*(3), 275-279.

Slavin, R. L. (1994). Effective Classroom, Effective Schools: A Research Base for Reform in Latin American Education. En: J. M. Puryear y J. J. Brunner (Eds.), *Education, Equity and Economic Competitiveness in the Americas* (págs. 7-28). Washington DC: Organization of American States.

Smith, M. S., Fuhrman, S. H., y O'Day, J. (1994). National Curriculum Standards: Are They Desirable and Feasible? En: R. F. Elmore y S. H. Fuhrman (Eds.), *The Governance of Curriculum* (págs. 13-29). Alexandria, VA: Association for Supervision and Curriculum Development.

Survey of Mathematics and Science Opportunities. (1992a). *TIMSS Mathematics Curriculum Framework* (38). East Lansing, MI: SMSO.

Survey of Mathematics and Science Opportunities. (1992b). *TIMSS Science Curriculum Framework* (37). East Lansing, MI: SMSO.

The World Bank. (1995). *Claiming the Future: Choosing Prosperity in the Middle East and North Africa*. Washington DC: The World Bank.

Valverde, G. A., y Schmidt, W. H. (1995). *An Exploratory Analysis of the Content and Expectations for Student Performance in Selected Mathematics and Biology School-Leaving Examinations from the Middle East and North Africa* . Washington DC: The World Bank.

Wiley, D. E. (1982). The Vicious and the Virtuous: ETS and College Admissions. *Contemporary Education Review, 1*(2), 85-101.

Zedillo Ponce de León, E. (1993). Presentación. En: Ministerio de Educación Pública (Ed.), *Educación Básica Secundaria: Plan y Programa de Estudios* (págs. 7). México D. F.: Ministerio de Educación Pública.

CAPÍTULO 4

SEGUIMIENTO INTERNACIONAL DE LAS METAS DE DESARROLLO HUMANO: EL CASO DE LA SECRETARÍA PRO TÉMPORE DE LAS AMÉRICAS

María Inés Cuadros Ferré

El interés de todos los países en mejorar la educación está motivando el surgimiento de acuerdos y consorcios sin precedentes. Este proceso conlleva comparaciones entre países, intercambio de metodologías y desarrollo de mecanismos de seguimiento. La experiencia del seguimiento de las metas establecidas para el desarrollo humano en el que se comprometieron los países de las Américas ilustra esta tendencia y brinda una oportunidad de analizar sus contribuciones y potencial, así como sus obstáculos y condiciones. Este documento sintetiza los principales acuerdos y metas suscritos por los gobiernos de las Américas en el campo de la niñez, a partir de la aprobación de la Convención Internacional sobre los Derechos del Niño. Describe el mecanismo de seguimiento regional a los acuerdos, creado por los gobiernos en 1994 y hoy conocido como Secretaría pro témpore; hace un recuento de la labor desarrollada durante su primer período de ejecución (1994-1996) y analiza sus logros y dificultades. Finalmente, el documento presenta algunas recomendaciones y propuestas para impulsar compromisos y establecer metas educativas de carácter internacional y reflexiona sobre el futuro de este tipo de mecanismos de seguimiento.

ANTECEDENTES

La Convención Internacional sobre los Derechos del Niño y la Cumbre Mundial en Favor de la Infancia

El 20 de noviembre de 1989 se aprobó en el seno de las Naciones Unidas, la Convención Internacional sobre los Derechos del Niño, la cual entró en vigor el 2 de septiembre de 1990. Su espíritu transforma al niño de objeto de una protección especial en sujeto de una amplia gama de derechos y libertades (en el ámbito de la

educación, ver Cuadro 1). Así mismo la Convención reconoce la dignidad del Niño como persona y en consecuencia, el cumplimiento de sus derechos se hace obligatorio y jurídicamente exigible. A principios de 1997 la Convencion había sido ratificada por 190 Estados[1].

Con el propósito de impulsar la Convención, y bajo el auspicio de las Naciones Unidas, el 30 de septiembre de 1990 se celebró en Nueva York la Cumbre Mundial en Favor de la Infancia. En ella 71 Jefes de Estado y de Gobierno y representantes de 88 países suscribieron la Declaración Mundial para la Supervivencia, el Desarrollo y la Protección de los Niños, comprometiéndose a promover la ratificación de la Convención de los Derechos del Niño a fin de convertirla en instrumento de validez legal en sus países y a realizar todos los esfuerzos posibles para alcanzar en el año 2000, mejores condiciones de vida y desarrollo para la niñez. La Declaración establece el compromiso de realizar un programa de 10 puntos centrales que se operacionalizan en un Plan de Acción Mundial con sus correspondientes metas. Este Plan define objetivos sectoriales de apoyo, agrupados en las temáticas de salud y educación de la mujer, nutrición, salud infantil, agua y saneamiento, educación básica y protección de niños y adolescentes en circunstancias especialmente difíciles.

Respecto a la educación, el Plan de Acción definió como dos de las metas más importantes para alcanzar en el año 2000 las siguientes: garantizar el "acceso universal a la educación básica y finalización de la enseñanza primaria por lo menos por el 80 por ciento de los niños en edad escolar y reducción de la tasa de analfabetismo de los adultos al menos a la mitad del nivel registrado en 1990, otorgando especial importancia a la alfabetización de la mujer." Para el cumplimiento de estas metas se establecieron cuatro objetivos específicos:

1. Ampliación de las actividades de desarrollo en la primera infancia, incluidas intervenciones apropiadas y de bajo costo con base en la familia y en la comunidad.

2. Acceso universal a la educación básica y finalización de la enseñanza primaria para al menos el 80 por ciento de los niños en edad escolar, mediante la enseñanza escolar o la educación no académica con un nivel de aprendizaje comparable, haciendo hincapié en reducir las disparidades actuales en la educación de niños y niñas.

3. Reducción de la tasa de analfabetismo entre los adultos a por lo menos el 50 por ciento del nivel que tenía en 1990, subrayando la alfabetización de las mujeres.

4. Mayor adquisición por parte de individuos y familias de los conocimientos, técnicas y valores necesarios para vivir mejor, que se les han de proporcionar a través de todos los cauces educativos, incluidos los medios de difusión y otras formas de comunicación y acción social modernas y tradicionales, cuya efectividad se mediría en función de los cambios logrados en el comportamiento.

Cuadro 1. Convención Internacional sobre los Derechos del Niño: derechos en el campo educativo

Artículo 28
1. Los Estados Partes reconocen el derecho del niño a la educación y, a fin de que se pueda ejercer progresivamente y en condiciones de igualdad de oportunidades ese derecho, deberán en particular: a) implantar la enseñanza primaria obligatoria y gratuita para todos; b) fomentar el desarrollo, en sus distintas formas, de la enseñanza secundaria, incluida la enseñanza general y profesional, hacer que todos los niños dispongan de ella y tengan acceso a ella y adoptar medidas apropiadas tales como la implantación de la enseñanza gratuita y la concesión de asistencia financiera en caso de necesidad; c) hacer la enseñanza superior accesible a todos, sobre la base de la capacidad, por cuantos medios sean apropiados; d) hacer que todos los niños dispongan de información y orientación en cuestiones educativas y profesionales y tengan acceso a ellas; y e) adoptar medidas para fomentar la asistencia regular a las escuelas y reducir las tasas de deserción escolar.
2. Los Estados Partes adoptarán cuantas medidas sean adecuadas para velar porque la disciplina escolar se administre de modo compatible con la dignidad humana del niño y de conformidad con la presente Convención.
3. Los Estados Partes fomentarán y alentarán la cooperación internacional en cuestiones de educación en particular, a fin de contribuir a eliminar la ignorancia y el analfabetismo en todo el mundo y de facilitar el acceso a los conocimientos técnicos y a los métodos modernos de enseñanza. A este respecto se tendrán especialmente en cuenta las necesidades de los países en desarrollo.

Articulo 29
1. Los Estados Partes convienen en que la educación del niño deberá estar encaminada a: a) desarrollar la personalidad, las aptitudes y la capacidad mental y física del niño hasta el máximo de sus posibilidades; b) inculcar al niño el respeto a los derechos humanos y las libertades fundamentales y de los principios consagrados en la Carta de las Naciones Unidas; c) inculcar al niño el respeto a sus padres, a su propia identidad cultural, a su idioma y sus valores, a los valores nacionales del país en que vive, del país del que sea originario y de las civilizaciones distintas de la suya; d) preparar al niño para asumir una vida responsable en una sociedad libre, con espíritu de comprensión, paz, tolerancia, igualdad de los sexos y amistad entre todos los pueblos, grupos étnicos, nacionales y religiosos y personas de origen indígena; e) inculcar al niño el respeto al medio ambiente natural.

La Convención y los Planes Nacionales de Acción en Favor de la Infancia en América

La Declaración de los Jefes de Estado produjo en América un impacto muy importante que se tradujo en la ratificación temprana de la Convención (hasta la fecha, 33 de los 34 Estados la han ratificado) y en la formulación de Planes Nacionales de Acción (PNAs) en favor de la infancia[2]. En el campo del desarrollo y la educación se observa la inclusión, en los PNAs, de los siguientes temas: (a) programas de estimulación temprana; (b) programas de educación inicial; (c) proyectos de escuelas interculturales; (d) fortalecimiento de la educación básica a nivel de coberturas, acceso y calidad; (e) definición de planes decenales de

educación; (f) proyectos de reformas educativas; y (g) sistemas de seguimiento del desarrollo.

Para hacer seguimiento y examinar los progresos realizados en el cumplimiento de las obligaciones contraídas por los Estados Parte en la Convención se creó el Comité de los Derechos del Niño, al cual se le deben presentar informes sobre las medidas que se hayan adoptado para dar efecto a los derechos reconocidos en la Convención y sobre el progreso que hayan realizado en cuanto al goce de estos derechos.

Todos los informes de países al Comité incluyen un capítulo sobre los avances y las dificultades en el campo educativo. Así mismo los informes de organismos no gubernamentales presentan en su mayoría análisis críticos sobre el cumplimiento del derecho a la educación. En este sentido su preocupación se concentra en el deterioro de la escuela pública en varios países, las fallas en la calidad de la educación, la alta deserción y repitencia, la escasa atención a la educación inicial de los niños en situación de mayor pobreza, las diferencias tan marcadas entre el acceso a la educación en las zonas urbanas en relación con las rurales, la falta de equidad entre los sexos, la falta de desarrollo de los proyectos pedagógicos donde los hay, el maltrato que reciben muchos niños y niñas en la escuela, la asociación entre niño trabajador y niño que deserta de la escuela y las dificultades que tienen los niños bilingües para acceder a la educación en sus lenguas maternas por fallas en los sistemas educativos.

Por otra parte, las naciones americanas, con el apoyo de los organismos de las Naciones Unidas, iniciaron en 1992 una serie de reuniones periódicas con el propósito de apoyarse en la formulación de los Planes Nacionales de Acción en Favor de la Infancia y para llevar a cabo un seguimiento de los logros obtenidos y las dificultades encontradas en su aplicación. La primera de estas reuniones se llevó a cabo en el Brasil. En abril de 1992, con el apoyo de la Organización Panamericana de la Salud (OPS/OMS), se invitó a representantes del sector salud y a algunos parlamentarios de los países americanos para que formularan o revisaran los Planes Nacionales de Acción de sus países. Posteriormente y por iniciativa del Gobierno de México se celebró una Reunión Ministerial Latinoamericana sobre Infancia y Política Social con el propósito de evaluar el avance en el cumplimiento de las metas y analizar aspectos tales como la coordinación intersectorial, la cooperación entre países y la financiación de los Planes. Al finalizar se firmó la Declaración de Tlatelolco, la cual reafirma el compromiso impostergable de los gobiernos de cumplir con los compromisos de la Cumbre Mundial.

Como una conjugación de los dos eventos anteriores, se celebró en abril de 1994, en Santafé de Bogotá, la Segunda Reunión Americana sobre Infancia y Política Social, la cual convocó a Ministros y representantes de 28 países de las Américas, responsables de los acuerdos sobre la niñez en la Región. Los temas centrales tratados durante la reunión versaron sobre: Institucionalización de la política social y los Planes Nacionales de Acción en favor de la Infancia, balance de las metas para

la infancia a la mitad de la década y desafíos, descentralización y municipalización, financiamiento y cooperación internacional y horizontal. El hito de la reunión fue la suscripción del Pacto de Nariño. A continuación se describe este acuerdo, se informa sobre el desarrollo, las acciones y el balance de la Secretaría creada como parte del mismo, hasta llegar al Acuerdo de Santiago, firmado en agosto de 1996 en Santiago de Chile con motivo de la Tercera Reunión Americana sobre Infancia y Política Social y vigente en la actualidad.

El Pacto de Nariño

Este acuerdo reafirma la obligación de dar cumplimiento a los acuerdos suscritos durante la Cumbre Mundial, declara que el eje de la política social y económica debe ser el desarrollo humano, señala el imperativo de invertir en la niñez para garantizar un desarrollo humano sustentable y equitativo y propone medidas a tomar por parte de los gobiernos para superar los factores estructurales de la pobreza y lograr la eficiencia y la productividad de la economía.

El Pacto está estructurado en seis secciones. La primera de ellas hace relación a las metas del desarrollo humano; la segunda a la institucionalización; la tercera se refiere a los compromisos en el campo de la descentralización, gestión pública y participación social; la cuarta a los propósitos de la financiación y cooperación internacionales; la quinta a los mecanismos de seguimiento regional y una última recoge las consideraciones finales. En el anexo se detalló la forma en que debía llevarse a la práctica el mecanismo regional de seguimiento.

En el campo educativo el acuerdo señala para los países americanos caminos claros a seguir. En lo referente a las metas, además de confirmar las definidas en la Cumbre Mundial, se destaca la necesidad de desarrollar una educación inicial basada en la comunidad y la familia, acordando identificar metodologías y estrategias para ampliar las coberturas y desarrollar los modelos alternativos de atención para el grupo de los más pequeños. Por otra parte, llama la atención sobre la calidad de la educación y, en este sentido, el compromiso suscrito es elevar la calidad de la educación primaria introduciendo reformas curriculares, aumentando la inversión en infraestructuras, textos apropiados, capacitación de maestros y asegurando que el período escolar sea adecuado a las necesidades de los países y permita un aprendizaje efectivo. Por último, identifica a la educación y capacitación técnica vocacional para los jóvenes como meta necesaria para los países americanos.

Más allá de las metas en el campo propiamente educativo, los acuerdos anteriormente mencionados son vitales para el desarrollo educativo. Así, por ejemplo, se resuelve apoyar los procesos de descentralización en los ámbitos regionales y municipales y se considera muy importante promover la participación de diversos grupos sociales en el cumplimiento de los Planes Nacionales de Acción, entre los que se encuentran los educativos. Igualmente se hace un llamado a continuar los esfuerzos para incrementar la inversión social en los países, asi como a mejorar la eficiencia, eficacia y equidad a fin de contribuir a garantizar los derechos

y las necesidades básicas de los más pobres del hemisferio.

LA SECRETARÍA PRO TÉMPORE AMERICANA: 1994- 1996[3]

Al crear el mecanismo de seguimiento, el Pacto inauguró la que es hoy Secretaría pro témpore de Seguimiento a los Acuerdos sobre la Niñez en las Américas. Las principales funciones que se le asignaron fueron las de: (a) promover el intercambio de experiencias exitosas de desarrollo social vinculadas con la infancia, (b) promover el intercambio intrarregional de información sobre metodologías e instrumentos de política social, (c) divulgar masivamente al público en general información relativa a los temas sociales especialmente los vinculados con la infancia y (d) apoyar la creación o el fortalecimiento de los sistemas de información nacionales o subregionales.

El gobierno del país sede de la última reunión americana sobre infancia y política social es el encargado de coordinar el mecanismo regional establecido, hasta la celebración de la siguiente reunión. En virtud de esta definición, el Gobierno de Colombia fue el encargado de iniciar la Secretaría pro témpore Americana. Esta Secretaría se instaló en el Departamento Nacional de Planeación y fue dirigida por la misma dependencia que tenía a su cargo la Coordinación del Plan Nacional en Favor de la Infancia. Su labor inicial consistió en definir su papel, sus actividades y la estrategia y mecanismo de relación con los países, las agencias de cooperación y los organismos no gubernamentales.

La Secretaría se definió como un mecanismo de apoyo establecido por los países con los propósitos y acciones previamente señalados por el Pacto de Nariño. Por su naturaleza de apoyo al seguimiento de los Planes de Acción en Favor de la Infancia, los cuales tienen su origen y fundamento en la Convención Internacional sobre los Derechos del Niño y en la Declaración Mundial para la Supervivencia, el Desarrollo y la Protección de los Niños del Mundo, la Secretaría consideró necesario realizar todas sus acciones bajo la óptica de tales mandatos.

Por otra parte, teniendo en cuenta que el UNICEF estaba ejerciendo la Secretaría del Comité Interagencial[4], se consideró que ésta sería la entidad con la cual se trabajaría directamente y a través de la cual se establecería la relación con las agencias y los organismos internacionales.

El seguimiento de las metas

Para iniciar su actividad, además de las labores de difusión del Pacto, se consideró tarea vital el que cada meta expresada en él tuviera expresión en uno o varios indicadores, de manera que con parámetros cuantitativos o cualitativos se lograra hacer, por parte de los países con el apoyo de la Secretaría, un verdadero seguimiento de lo acordado. Para ello, se elaboró una tabla de indicadores con sus correspondientes definiciones y fórmulas para obtenerlos. Posteriormente se solicitó a los países diligenciar un cuestionario con información referida a las metas,

los indicadores sugeridos, la información disponible más reciente, la fuente para obtenerla, la periodicidad de los datos, el nivel de desagregación posible y el valor de la meta por país para los años 1995 y 2000.

Venticuatro países del hemisferio diligenciaron la encuesta. El análisis realizado por la Secretaría permitió identificar los logros y las dificultades de los países en su conjunto en el campo de los sistemas de información necesarios para supervisar las metas sobre niñez. Los resultados de la encuesta mostraron que los indicadores del desarrollo de la educación denominados tradicionales, tales como las tasas de escolaridad, deserción y el número de profesores, estaban disponibles en la región. Pero los nuevos indicadores, referidos a la calidad de la educación y a los resultados del sistema, se hallaban excluidos de los sistemas de información. No existían indicadores para medir, por ejemplo, la adquisición de habilidades, el número de horas de aprendizaje y la inversión en textos escolares. Sin estos indicadores, no es posible monitorear los resultados de los programas de reforma de la educación en los que se encuentran empeñados la mayoría de los países.

A la fecha de la encuesta (primer semestre de 1995), las fuentes de datos para las metas educativas eran principalmente los Ministerios de Educación Nacional u Oficinas de Planeación Educativa. Figuraban también los departamentos o institutos de estadística de los países, el sistema de información educativa y la UNESCO. El principal instrumento utilizado era el registro administrativo, además de los censos, las encuestas permanentes y las específicas.

La desagregación subnacional estaba bastante generalizada en la Región, pero la de género y edad era muy incipiente. Se consideró que esta falla podía deberse a que los instrumentos utilizados no eran los adecuados, a un procesamiento deficiente de los cuestionarios o a que no se estuvieran aprovechando todos los datos disponibles al momento de responder el cuestionario. Cualquiera que sea el caso, debe prestarse mucha atención a este hallazgo porque los análisis generacionales son mucho más ilustrativos para la planificación del sector y también porque la equidad hace necesario el análisis por género.

Las conclusiones generales señalaron que (1) los indicadores de salud y educación tenían una mayor disponibilidad de información que los de derechos cívicos; (2) en todas las áreas y aún en indicadores considerados tradicionales se observó una ausencia de estandarización que impide la comparabilidad entre países; (3) la mayor falta de estandarización se encontró en los rangos de edad aplicable, los factores de ampliación de las tasas y las categorías incluidas; (4) en general, se carecía de metas para los distintos indicadores, así los países hubieran ratificado los acuerdos internacionales o no; y (5) no existe una idea clara de lo que es y de cómo debe funcionar un sistema de información para el seguimiento de las metas.

En las recomendaciones generales sobresalió la necesidad de crear conciencia en las instituciones sobre lo que significa el seguimiento de las metas y su utilidad para cada país y para el conjunto de ellos. Igualmente se identificó como prioridad el

continuar con el diseño y la estandarización de indicadores, fortalecer las instituciones en el procesamiento de la información con el fin de que se aprovechara mejor el potencial de las distintas fuentes y apoyar a los países en el establecimiento de verdaderos sistemas de información.

La Secretaría pro témpore, conjuntamente con el Gobierno de Guatemala y el UNICEF, preparó y llevó a cabo una reunión técnica en agosto de 1995. Ésta tuvo como propósito analizar los mecanismos de seguimiento de las metas de los Planes Nacionales de Acción, en lo referente a las fuentes de información, la pertinencia de los indicadores propuestos y la frecuencia y regularidad con que se deben evaluar los avances y logros, e intercambiar experiencias exitosas en el área de los sistemas de información entre los países de la Región.

Las ponencias de esta reunión mostraron cómo las Encuestas de Hogares y la de Indicadores Multiples constituyen valiosos instrumentos; señalaron algunos de los desafíos en la medición de insumos, procesos y resultados en educación; informaron sobre el desarrollo, uso y presentación de bases de datos y enfatizaron la importancia del uso de la información educativa para la toma de decisiones políticas.

Los participantes a su vez identificaron la necesidad de generar indicadores educativos con rigurosidad e imaginación y de crear indicadores para medir el impacto de los logros educativos en la disminución de la pobreza. Igualmente señalaron las dificultades que existen para establecer cobertura y metodologías en la educación preescolar, en especial por su carácter privado en la mayoría de los países, y el desafío que supone el conocer la situación educativa de aquellos niños y niñas que no se encuentran vinculados al sistema formal.

Entre las recomendaciones que se desarrollaron al término de la reunión cabe destacar las siguientes: (1) fortalecer los sistemas de información a fin de que proporcionen los insumos necesarios para realizar evaluaciones periódicas del avance hacia las metas trazadas en los programas y cuyos resultados faciliten la toma de decisiones para generar acciones y orientar el gasto social de acuerdo con las necesidades reales de la población; (2) mejorar y normalizar la disponibilidad, confiabilidad, oportunidad y utilización de la información referida a las condiciones de la niñez; (3) fortalecer sus sistemas de información social, potenciar la producción, el análisis y la utilización tanto a nivel nacional y regional, como local y analizar la situación social según sexo, grupo étnico y población en riesgo; (4) enfatizar la participación de la sociedad civil en la producción, el análisis y la utilización de la información; y (5) recabar y difundir información para mejorar la capacidad de las niñas, niños y jóvenes de participar, expresarse libremente y recibir e impartir información.

AVANCES EN EL LOGRO DE LAS METAS DE DESARROLLO HUMANO

Al finalizar el período bajo la responsabilidad del Gobierno de Colombia, la Secretaría pro témpore realizó un análisis de las actividades llevadas a cabo hasta entonces por las naciones americanas para el cumplimiento de las metas. Para elaborar el informe se revisaron las respuestas de los países a un cuestionario evaluativo diseñado por la Secretaría, los balances de mitad de década entregados por los países al UNICEF, los informes gubernamentales y los no gubernamentales presentados por los países al Comité de los Derechos del Niño, los informes especializados disponibles en las agencias de las Naciones Unidas y los propios documentos de la Secretaría.

Crecimiento y desarrollo

Los informes de los países señalaron dos modalidades de atención para los niños y niñas entre cero y cinco años :

- *Programas de vigilancia*: adelantados generalmente por el sector salud, en los cuales se realizan actividades de seguimiento del crecimiento y desarrollo. En sólo dos países[5] existía una escala validada para medir de manera sistemática el crecimiento y desarrollo de niños y niñas.

- *Programas de cuidado diario*: en éstos se identificaron por lo menos tres modalidades, a saber, (1) Hogares comunitarios y Hogares de cuidado diario, en los cuales se da una alta participación comunitaria; (2) redes de atención a partir de la asociación de esfuerzos institucionales de los gobiernos, con iniciativas locales independientes y de organizaciones no gubernamentales. Estas dos modalidades desarrollan componentes de nutrición y protección y generalmente reciben algún tipo de subsidio estatal; y (3) programas de atención preescolar, diseñados principalmente para las clases medias y altas y organizados por el sector privado.

La demanda de atención para estos niños y niñas presenta un incremento continuo debido a que la mayoría de las madres trabajan y requieren de instituciones u organizaciones que contribuyan con el cuidado y la crianza de sus hijos. Los países de la región identificaron como prioridad la capacitación y cualificación de los padres de familia así como la del personal responsable del cuidado de los niños y niñas más pequeños, para apoyar el desarrollo infantil desde los primeros meses de vida, buscando recuperar la función educadora y protectora de la familia y de la comunidad.

Educación inicial, preescolar y aprestamiento básico

Se observa en la Región una tendencia a fortalecer programas de educación inicial,

evidenciándose la expansión y diversificación de programas y metodologías basados en esquemas no formales con una alta participación comunitaria. Los países resaltan los beneficios de promover estos programas por el impacto en el desarrollo de los niños y niñas, la mayor facilidad que les da para ingresar al sistema educativo formal y la reducción de los índices de repitencia. No obstante, sólo el 20 por ciento de los niños y niñas de cinco y seis años tiene acceso a este tipo de programas[6].

Educación básica

Los países de América Latina y el Caribe registran simultáneamente altos niveles de matrícula primaria y elevados índices de repitencia y deserción. De esta forma el acceso neto a la educación básica primaria es del 86 por ciento para los niños y niñas entre seis y once años de edad, hallándose el mínimo (75 por ciento) en los países centroamericanos y el máximo (93 por ciento) en el Caribe de habla inglesa. El 42 por ciento de quienes ingresan en la escuela repite el primer grado y el 30 por ciento repite el segundo. El promedio de repitencia para todos los grados de primaria es del 30 por ciento, con costos que ascienden a la cifra de tres mil quinientos millones de dólares americanos.

Las causas asociadas al hecho de que 22 millones de niñas y niños de la Región repitan grados escolares son la baja calidad del sistema educativo, la poca disponibilidad de recursos de las escuelas, la carencia de ayudas pedagógicas y la falta de capacitación de los maestros. También son causa las propias características de niños y niñas a nivel de desarrollo, nutrición, adaptación y desarrollo psicológico. Adicionalmente, los logros de una buena parte de los estudiantes que permanecen en el sistema son significativamente inferiores a los alcanzados en otras regiones del mundo en desarrollo. Los mejores desempeños se han observado en el Caribe de habla inglesa y en los estudiantes de colegios privados del resto de los países de la zona.

Elevar la calidad de la educación es uno de los retos de la Región, lo cual implica, entre otros esfuerzos: enfatizar el aprendizaje de habilidades básicas para la lectura, escritura y aritmética; promover el desarrollo de materiales y guías que respondan a la diversidad de los grupos de estudiantes; ofrecer educación más apropiada y de mejor calidad en las zonas rurales y a los grupos étnicos; reforzar la capacidad de gestión de las entidades territoriales y garantizar la capacitación, la adecuación de las condiciones salariales y el aporte de incentivos para los maestros.

Educación secundaria

En esta etapa se ponen de manifiesto con mayor severidad los efectos de la deserción y la falta de accesibilidad al sistema para la gran mayoría de los adolescentes. Aunque la cobertura aumentó del 14 por ciento en 1960 al 55 por ciento en 1993, casi la mitad de los adolescentes del continente continúa sin alcanzar este nivel educativo.

Las implicaciones de las bajas coberturas son variadas y afectan significativamente el desarrollo de los países. Por una parte, los procesos de la economía demandan un capital humano cada vez más cualificado y, por otra, el bajo nivel educativo trae como consecuencia ciertos efectos negativos de tipo intergeneracional tales como familias constituidas más tempranamente, más numerosas y con menor capacidad para generar recursos, elevando la probabilidad de permanecer o caer en la pobreza y en la miseria. La Región tiene que realizar un esfuerzo para encontrar alternativas para los adolescentes, aumentando las coberturas, mejorando la calidad y diversificando las oportunidades de formación ocupacional una vez concluida la educación basica.

El potencial del sector educativo

La Secretaría llamó la atención de los países sobre la alta rentabilidad de una inversión oportuna en educación, su alta tasa de retorno (estimada en un 27 por ciento) y su repercusión no sólo en el campo de la producción sino en los niveles de salud y calidad de vida de la población.

Igualmente, considerando la circunstancia de que las instituciones educativas que mantienen una población "cautiva"—lo cual constituye una oportunidad única para conocer las condiciones de vida, su entorno familiar y social, sus ocupaciones y su nivel de salud y así tener una visión integral de cada alumno y alumna—hizo un llamado para que el sector tenga un papel más activo en este conocimiento, se fortalezcan sus alianzas y coordinación de acciones con otros sectores y se diseñen mecanismos para asesorar, orientar, detectar y referir a los alumnos y sus famlias con base en las condiciones particulares que presenten.

EL ACUERDO DE SANTIAGO

En agosto de 1996, el Gobierno de Chile, con el apoyo de la Secretaría pro témpore y el UNICEF, convocó la Tercera Reunión Ministerial Americana sobre Infancia y Política Social. Los objetivos de la reunión fueron establecidos conjuntamente entre las partes organizadoras de la misma y se definieron para: (1) analizar el avance alcanzado a mitad de la década en el cumplimiento de los compromisos suscritos en la Cumbre Mundial en Favor de la Infancia y el Pacto de Nariño; (2) reafirmar el compromiso de los países del hemisferio de mejorar las condiciones de vida de la infancia, mediante la revisión de las metas a nivel regional para el año 2000 y más allá establecer nuevos retos y estrategias; (3) fomentar el intercambio de experiencias entre los países de la Región y con las agencias de cooperación internacional y fortalecer los mecanismos de cooperación horizontal.

La presentación realizada por la Secretaría sobre el balance de la experiencia hasta el momento, sirvió de base para las discusiones de los delegados sobre logros, obstáculos y retos enfrentados por los países durante el proceso. Al finalizar la reunión los países suscribieron el Acuerdo de Santiago en el cual reafirmaron su

compromiso para:

- Desarrollar, en el marco de la Convención Internacional de Derechos del Niño, una política sistémica, integral y progresiva sustentada en los derechos y las responsabilidades de todos los actores, el desarrollo de sujetos sociales—y no de objetos de tratamiento de programas—, la participación eficiente en la formulación de políticas nacionales y la producción y distribución de bienes y servicios sociales.
- Prevenir y enfrentar problemas emergentes tales como el SIDA, las adicciones, el embarazo precoz, el abandono, el abuso sexual, el trabajo infantil y la violencia.
- Asegurar el cumplimiento universal y la sostenibilidad de las metas de bienestar de niños, niñas y adolescentes para el año 2000, mediante la realización de importantes esfuerzos adicionales para acelerar, expandir y aumentar la eficiencia, eficacia y equidad de la política social.
- Promover y respaldar los programas dirigidos a apoyar la familia.
- Dar un verdadero impulso a la inversión en el desarrollo de recursos humanos, tendente a elevar y mantener la calidad de vida de los pueblos del hemisferio.
- Fortalecer los procesos de democratización y la consolidación de una cultura de cooperación en las generaciones jóvenes y la educación en hábitos y actitudes de solidaridad, valores cívicos y derechos humanos.
- Fortalecer los sistemas de información social basados en mecanismos periódicos de recolección de información, incluir los principales indicadores sociales y fomentar su utilización en la gerencia social, mejorando su disponibilidad, confiabilidad, oportunidad, comparabilidad y desagregación.

Específicamente en el ámbito educativo, el Acuerdo establece tres metas principales derivadas de la Cumbre Mundial pero adaptadas a la Región y 14 metas auxiliares. Las metas principales son las siguientes: la primera, lograr el acceso universal a la educación primaria con equidad, desde la perspectiva de género, geográfica, étnica, de niveles socioeconómicos y de grupos de necesidades especiales; la segunda, aumentar a más del 80 y 70 por ciento el porcentaje de niños y niñas que terminan cuarto grado y primaria, respectivamente y aumentar significativamente el acceso a la educación secundaria; y la tercera, reducir a la mitad la tasa de analfabetismo de los adultos, con especial atención a la alfabetización de la mujer.

Por su parte, las metas auxiliares indicadas en el Acuerdo son:

- Establecer sistemas de información sobre educación inicial y educación de los padres que permitan el adecuado seguimiento y la evaluación de los programas en esos campos.
- Legislar, establecer políticas y regulaciones para mejorar la educación inicial/ preescolar y aumentar las asignaciones presupuestarias para este sector..
- Promover la extensión y el mejoramiento de programas de educación inicial/ preescolar basados en la familia y en la comunidad.
- Promover la extensión de programas de educación de los padres el sobre el

desarrollo infantil, haciendo énfasis en las responsabilidades del padre en la crianza.

- Reducir a la mitad las tasas de repitencia en los primeros grados de la primaria.
- Identificar estrategias para disminuir la deserción escolar.
- Elevar los niveles de comprensión lectora, de escritura, de cálculo y el número de los niños y niñas que terminan cuarto grado.
- Elevar la equidad y calidad de la educación básica, introduciendo transformaciones en: (1) el plano de la gestión, fortaleciendo la administración, planificación, ejecución y supervisión del proceso educativo, promoviendo una mayor participación de los alumnos, los padres y la comunidad; (2) el plano curricular, adecuando a las necesidades de la persona, la comunidad y la sociedad, promoviendo metodologías participativas individuales y grupales centradas en el educando, ampliando el papel del profesor como facilitador del proceso de aprendizaje y proporcionando a los alumnos materiales adecuados; y (3) el plano de las inversiones, con especial atención a la equidad, mediante asignación de recursos suficientes de infraestructura, dotación y desarrollo profesional de maestros, textos y duración de la jornada escolar.
- Asegurar la organización y el uso de sistemas de medición del nivel y la calidad del aprendizaje.
- Desarrollar opciones flexibles de educación y capacitación técnica vocacional para los jóvenes, especialmente cursos de informática y empresariales.
- Incorporar en el currículo, en los materiales y métodos pedagógicos usados en todos los establecimientos de educación preescolar, primaria y secundaria educación sobre derechos humanos comenzando con los derechos del niño, la niña y la mujer; y destrezas para la vida, el medio ambiente, la nutrición y la salud, incluyendo salud reproductiva.
- Revisar y detectar sesgos de género y cualquier otro tipo de discriminación en todos los materiales y métodos de enseñanza preescolar, primaria y secundaria, tomar medidas correctivas; y asegurar la inclusión de las nociones de equidad de género y social en los nuevos materiales y metodologías de enseñanza.
- Ejecutar y expandir programas y acciones para la promoción del estado de las niñas y las adolescentes y del respeto mutuo entre niños y niñas en las instituciones de educación preescolar, primaria y secundaria.
- Desarrollar programas de rehabilitación para apoyar a los niños con discapacidades y a sus familias.

CONTRIBUCIÓN DE LA SECRETARÍA PRO TÉMPORE

Varias fueron las lecciones aprendidas durante los dos primeros años de operación de la Secretaría pro témpore. La mayoría de los logros directos de una red de intercambio y seguimiento como la promovida por la Secretaría pro témpore, son de carácter general, simbólico y político. La Secretaría logró mantener el interés y el foco de la política social en los niños, alimentar el proceso de definición de las políticas nacionales y concientizar a los gobiernos sobre la necesidad de evaluar sus programas de desarrollo humano y mejorar sus sistemas de información. La

Secretaría, además, convocó a gobiernos, agencias internacionales y organizaciones académicas en torno de la evaluación de objetivos de interés común.

Foco de atención nacional en la niñez

La Secretaría ha constituído un espacio definido y reconocido por los países que les ha permitido mantener la atención en los compromisos adquiridos por los gobiernos del hemisferio con la niñez, a pesar de los cambios de administración ocurridos en cada uno de ellos. Durante el período 1994-1996 más de quince países tuvieron cambio de Jefe de Estado, lo cual trajo consigo modificaciones de los Planes Nacionales de Desarrollo de los diversos países, cambios en la dirección de las políticas sociales y relevos en los cuadros directivos de las instituciones gubernamentales. La Secretaría realiza el seguimiento a los acuerdos sobre niñez, actúa como ilustradora de los compromisos, al tiempo que colabora para que se continúe focalizando la política social en la garantía, defensa y promoción de los derechos de la niñez[7].

En este sentido puede decirse que la Secretaría acompaña a los diseñadores y ejecutores de políticas, promoviendo que sus decisiones y acciones contribuyan a hacer realidad los derechos de la infancia y generen beneficios a mediano y largo plazos para niñas y niños.

Legitimación de compromisos y metas al más alto nivel

La promoción de reuniones ministeriales impulsadas por los países con el apoyo y asesoría de la Secretaría pro témpore y de las agencias internacionales constituye una oportunidad para centrar la atención de los gobiernos en el cumplimiento de sus compromisos con la niñez en sus países. Al mismo tiempo la suscripción de acuerdos a nivel regional permite revisar, precisar y legitimar al más alto nivel político las metas definidas.

Adaptación de metas

Las metas definidas en la Cumbre Mundial en favor de la Infancia son, por su naturaleza universal, indicativas para el hemisferio. Las reuniones efectuadas por los países americanos, con el apoyo de la Secretaría y de las agencias, han permitido precisar las metas, e incluso definir nuevas metas para algunos subgrupos de edad que no fueron contempladas en los acuerdos mundiales.

En el caso de la educación se observa un impulso a la educación inicial y a la educación flexible y técnica para los jóvenes, así como a la equidad y la calidad de la educación básica, más allá de la cobertura misma[8].

Mejoramiento y utilización de la información

La reunión de naturaleza técnica, impulsada por la Secretaría, señaló a los países la

importancia del seguimiento de las metas al tiempo que mostró cómo los sistemas de información desarrollados por los países no respondían con precisión a la formulación de los Planes Nacionales de Acción en Favor de la Infancia. Esto limitaba las posibilidades de evaluar en forma más objetiva el desempeño en materia social y de niñez, tanto a nivel de cada país como de la región en su conjunto.

La revisión de los sistemas de información social que se relacionaban con los programas y las actividades propios de estos Planes puso de manifiesto las múltiples dificultades en la producción de datos y su análisis y, sobre todo, en la utilización real que se le da a la información. Así mismo, se señaló la dispersión de fuentes e indicadores como una gran dificultad para consolidar la información requerida para el seguimiento a los compromisos.

Por otra parte, se reafirmó en los países la necesidad de disponer de la información para diseñar las políticas más acertadas para cada país, para focalizar las acciones con mayor efectividad, reforzar o disminuir la intensidad de las intervenciones específicas, obtener y asignar los recursos financieros necesarios y para conocer los avances, logros y obstáculos en las metas propuestas.

Se señaló al proceso de descentralización como una coyuntura importante para canalizar los esfuerzos nacionales tendentes a mejorar los registros e involucrar a los actores sociales en la generación y utilización de la información. También se mostró que perfeccionar los mecanismos de información requiere de la cooperación horizontal entre países y del apoyo técnico internacional de agencias y cooperantes.

Concientización sobre los problemas de desarrollo y protección de niños y niñas

La Convención Internacional sobre los Derechos del Niño incluye los derechos de todos los niños y niñas. El Plan de Acción, sin embargo, se centra en los problemas de supervivencia de la niñez. En el hemisferio, si bien subsisten graves problemas en esta última área, no cabe duda de que deben impulsarse otros derechos. El papel de la Secretaría y de las agencias de cooperación internacional ha sido el de ayudar a los países a definir cuáles de estos derechos deben ser priorizados. De esta forma se ha ido perfilando cada vez más la línea correspondiente al derecho al desarrollo y a la protección.

Análisis sobre temas relativos a la niñez

El balance que la Secretaría presentó durante la Reunión de Santiago es clara muestra de las posibilidades que ofrece un mecanismo como este. La posibilidad de acceder oportunamente a la información disponible en las entidades de gobierno de los distintos países, en las agencias de cooperación internacional y en otras fuentes, permitió organizar una presentación y documento que retroalimenta a los países al tiempo que les señala dónde se han obtenido los mayores logros y dificultades y sugiere hacia dónde se deben dirigir los esfuerzos conjuntos.

El balance identificó también los temas que se encontraban poco desarrollados, los desvíos y los temas emergentes[9]. Sobre éstos recomendó a los países y las agencias que concentraran parte de sus esfuerzos técnicos y financieros.

Intercambio de experiencias y metodologías exitosas

Una de las actividades que realizó la Secretaría, principalmente como parte de preparación de las reuniones y también durante las mismas, fue identificar experiencias y metodologías exitosas que los países habían realizado en distintas áreas relacionadas con la niñez. El intercambio de éstas constituye frecuentemente un aporte para los diseñadores de política y para los ejecutores de programas, ya que conocer cómo lo han hecho otros, les ayuda a visualizar las posibilidades que tienen de aplicarlas en un determinado país o región.

Existen algunas ventajas para el país que alberga una Secretaría de este tipo. En el caso de Colombia, por ejemplo, fue el fortalecimiento de la capacidad profesional y mantener un interés por la niñez en los altos espacios gubernamentales del país, en especial entre los planificadores. Adicionalmente, la relación constante y cercana con las agencias de cooperación internacional fortalece los vínculos del país con ellas. Un beneficio adicional es la formación que da a quienes tienen la responsabilidad de orientarla y coordinarla. Una mirada de conjunto siempre amplía el horizonte, enseña sobre temáticas antes no tratadas y garantiza la actualización tanto en la situación como en las respuestas dirigidas a garantizar los derechos de la niñez.

La Secretaría y sus relaciones con la comunidad internacional

En su carácter de ente constituido por la voluntad de los gobiernos de los países americanos, la Secretaría tiene una función esencial de seguimiento. Como coordinadora, está encargada de buscar el apoyo de organismos e instituciones de cooperación internacional para facilitar su funcionamiento. Igualmente, puede establecer canales de comunicación y coordinación con organismos o instancias intrarregionales o internacionales que trabajen en las áreas de infancia y desarrollo social.

Durante su primer período de funcionamiento, la Secretaría estableció relación con las agencias de cooperación internacional a través del UNICEF, respetando su calidad de coordinador interagencial. En la actualidad esta tarea es ejercida por el Fondo de Población de las Naciones Unidas (FNUAP).

La relación con el UNICEF fue extremadamente fructífera. Este organismo tiene una importante infraestructura a nivel regional y de los países, que puso a disposición de la Secretaría. De esta forma el trabajo fue mucho más eficiente y no se redoblaron esfuerzos que habrían sido por demás de un elevado costo para cualquier gobierno.

A partir de la experiencia colombiana, los gobiernos decidieron dar a la Secretaría la tarea de fortalecer las relaciones con el Comité Coordinador Interagencial y su vínculo con el UNICEF, como agencia especializada en temas de niñez y adolescencia.

Limitaciones y obstáculos en el seguimiento a los acuerdos y las metas

Como se mencionó anteriormente, el principal obstáculo para el seguimiento de los acuerdos y las metas en favor de la niñez lo constituye la falta de verdaderos sistemas de información social en la mayoría de los países. Existe gran heterogeneidad en la definición de indicadores y falta de información para documentar muchas de las metas.

Otro tipo de obstáculos derivan de la gran cantidad de cambios en los cuadros directivos y técnicos de los países. Con demasiada frecuencia hay que volver a informar sobre los acuerdos y las metas de orden internacional. Si bien las reuniones ministeriales han sido de un alto nivel político, en términos generales los representantes son del área social de los países (Ministros de Salud, Educación, Familia, Trabajo y Desarrollo Humano) y el cumplimiento de las metas en los países no depende sólo de ellos. En este sentido convendría incorporar más a los Ministros responsables de las áreas de Planificación y Finanzas. Una limitación que tiene el mecanismo es que no ha incorporado en el seguimiento a actores clave de la sociedad civil, entre los cuales las niñas y niños deberían tener un papel protagónico.

Factores clave a tener en cuenta en mecanismos internacionales semejantes

Una Secretaría pro témpore necesita el respaldo político de los países para que pueda operar. La tarea de seguimiento de los acuerdos sobre la niñez en la Región que se le ha dado a la actual Secretaría tiene su orígen y fortalecimiento a través del Pacto de Nariño y el Acuerdo de Santiago. Este es un requisito indispensable sin el cual difícilmente podría operar. El país que la asuma debe disponer de un personal cualificado que se encargue de dinamizarla. Este grupo requiere, además, de capacidad técnica, contactos con la comunidad académica y los organismos de decisión política. Independiente del personal se requiere de un área física y de un apoyo logistico importante para la comunicación y los viajes internacionales y las actividades de difusión.

El conocimiento, la información disponible y la infraestructura de una agencia internacional de cooperación, como el UNICEF para el caso de la niñez, es indispensable para que la Secretaría opere adecuada y eficientemente. De lo contrario, se pueden redoblar esfuerzos, ir por caminos paralelos o en contravía y perder la posibilidad de acercarse a los países y a las agencias.

La definición de compromisos y metas debe trabajarse ampliamente antes de llegar

a los acuerdos ministeriales. Así, hay que lograr previamente el consenso de los técnicos tanto de los gobiernos como de las agencias sobre la viabilidad de las metas. Igualmente, someter a la vía diplomática los proyectos de acuerdo, de lo contrario muchos países se abstienen de suscribirlos.

Este proceso de discusión continua permite un aprendizaje social que se manifiesta, por ejemplo, en una formulación de las metas de una manera cada vez más precisa, lo que facilita su seguimiento y evaluación. Esto se nota al comparar las formulaciones iniciales de las metas pretendidas con las aprobadas en Santiago. No obstante, las nuevas metas auxiliares no son fáciles de cumplir. Para lograrlas los países van a requerir de diversos apoyos y de una mayor eficiencia de sus propios sistemas educativos. En este sentido se recomienda a las organizaciones de cooperación internacional conocer a fondo las metas y chequear con los países su situación frente a ellas a fin de priorizar la asistencia técnica y financiera.

De igual forma que se han realizado reuniones específicas para revisar temas de niñez de los acuerdos, se podría pensar en una reunión específica, impulsada por la Secretaría con apoyo de las agencias especializadas en educación y niñez, de los responsables de la formulación de políticas y ejecución de programas educativos, que permita analizar los logros y obstáculos y definir las alternativas para promover el cumplimiento de las metas propuestas.

Se debe promover, también, la participación de otros actores como directores de escuela, profesores y padres de familia. Ninguna de las metas se puede cumplir si estos actores no se involucran, participan y deciden. Las organizaciones no gubernamentales internacionales y nacionales pueden desempeñar un papel protagónico en la coordinación de actores de la sociedad civil, tanto para la toma de decisiones, el apoyo técnico y la contribución con modelos exitosos, como para la realización de las tareas y el cumplimiento de las metas.

La participación de las niñas y niños, escolarizados y no escolarizados, merece mención especial. Ellos tienen derecho a la educación y a participar de una manera auténtica, no simbólica, de las decisiones que los afectan. Por ello la principal recomendación para evaluar las metas es incluir las opiniones de las niñas y los niños en la evaluación de las metas y, por supuesto, en la formulación y ejecución de las políticas y los programas.

Futuro del mecanismo de seguimiento a los compromisos americanos con la niñez

La Secretaría pro témpore continuará existiendo en cuanto los países así lo decidan. En la opinión de la autora de este documento, esta decisión estará basada en la utilidad que los gobiernos le encuentren. Cada país que la asuma le dará una dirección propia, conservando el mandato fundamental que es el de promover el intercambio de experiencias exitosas, definir una agenda regional de cooperación horizontal, apoyar y fortalecer los sistemas de información y evaluar

periódicamente el cumplimiento de las metas.

Probablemente una de las formas más efectivas para fortalecer la Secretaría es que ésta se apoye en alianzas y coordinadoras de organizaciones y grupos que compartan la filosofía de la Convención Internacional sobre los Derechos del Niño. Este tipo de organizaciones—que cuentan con gran experiencia y autoridad en el tema de la niñez—pueden ser fundamentales para la participación de la sociedad civil, incluidos los propios niños y niñas, en la praxis y el seguimiento de los derechos de la niñez y de las metas definidas por los gobiernos para alcanzarlos.

Así mismo, una fuerte coordinación con el Comité de los Derechos del Niño y con las Agencias de las Naciones Unidas y con la OEA a manera de consorcio internacional asegurará el seguimiento a la Convención, punto de partida de los acuerdos.

NOTAS

[1] Las naciones que aún no han ratificado la Convención son Somalia, las Islas Cook y los Estados Unidos de América.

[2] A finales de 1996, 23 de los 34 Estados habían definido los correspondientes PNAs, nueve estaban en proceso de elaboración y en dos países se había establecido como un documento de trabajo sin otro referente a nivel de políticas generales.

[3] Las acciones de la Secretaría pro témpore sólo son analizadas para el período 1994-1996 en razón a que la Secretaría fue entregada por el Gobierno de Colombia al Gobierno de Chile en agosto de 1996 y la autora de este documento no disponía de suficiente información sobre las labores emprendidas a partir de ese momento.

[4] El Comité está conformado por las siguientes agencias: Organización Panamericana de la Salud (OPS/OMS), Organización de las Naciones Unidas para la Agricultura y la Alimentación (FAO), Fondo de Población de las Naciones Unidas (FNUAP), Fondo de las Naciones Unidas para la Infancia (UNICEF), Banco Mundial (BM), Banco Interamericano de Desarrollo (BID), Organización Internacional del Trabajo (OIT), Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO), Agencia para el Desarrollo Internacional de los Estados Unidos (USAID), Programa de las Naciones Unidas para el Desarrollo (PNUD).

[5] Los países que reportaron tener una escala validada fueron Chile y Colombia. Algunos informaron estarla construyendo pero no tenerla en funcionamiento, concretamente el Perú y Costa Rica.

[6] De acuerdo con la información de UNESCO-OREAL, la atención en educación pre-primaria de la población de 0 a 5 años de edad llegaba a un 14 por ciento en 1989.

[7] La forma en que la Secretaría realizó esta labor principalmente fue con comunicaciones

informativas, enviando paquetes que contenían los Acuerdos a los nuevos gobiernos, solicitando información sobre quién estaba a cargo del seguimiento de los compromisos y requiriendo avances en el cumplimiento de los mismos. El elevado número de países que respondió las encuestas de la Secretaría, así como su participación en las Reuniones convocadas por ella confirma esta función.

[8] Una revisión de los distintos acuerdos suscritos permite inferir esta contribución de la Secretaría. Debe sin embargo aclararse que la Secretaría no realizó esta labor de manera independiente, sino acompañada por las Agencias especializadas de las Naciones Unidas y los propios países. La adaptación de metas no significa que los países empiecen a cumplirlas de inmediato, pero sí que empiecen a pensarlas y a tenerlas en cuenta al momento de definir políticas y programas nacionales y locales. Puede comprobarse esta afirmación al revisar los distintos Planes Nacionales de Desarrollo de los países. Ejemplo de ello son la inclusión en ellos de programas específicos para aumentar la cobertura de la educación inicial, para prevenir el maltrato infantil, atender la discapacidad, erradicar el trabajo infantil y modificar la legislación haciéndola acorde con la Convención Internacional.

[9] Los temas identificados en el Balance como emergentes fueron HVI/SIDA, violencia contra la niñez, trabajo y explotación infantil y salud reproductiva en la adolescencia.

Parte II
LECCIONES DE LA HISTORIA

CAPÍTULO 5

IMPACTO SOCIAL DE LOS SISTEMAS DE EVALUACIÓN DEL RENDIMIENTO ESCOLAR: EL CASO DE CHILE

Erika Himmel

La trayectoria de los sistemas de evaluación del rendimiento escolar en Chile proporciona lecciones de gran valor a los países de las Américas que están ejecutando programas nacionales de pruebas, al menos en dos aspectos relacionados directamente con la política: el de la responsabilidad institucional y el del impacto de los sistemas de evaluación en la sociedad. El Ministerio de Educación, las universidades y los centros de investigación han trabajado de manera coordinada para fortalecer la capacidad institucional y de recursos humanos del país para el desarrollo y la administración de diversos sistemas de evaluación. Este capítulo, además de describir el funcionamiento de tales sistemas, presenta un marco de referencia para la discusión del problema de la utilización de la información proporcionada por los sistemas de evaluación, dentro del cual se analiza el caso de Chile. Finalmente, la autora identifica los factores críticos relacionados con el uso de los resultados de la evaluación del rendimiento escolar.

INTRODUCCIÓN

Este capítulo, además de describir el funcionamiento de tales sistemas, presenta un marco de referencia para la discusión del problema de utilización de la información proporcionada por los sistemas de evaluación, dentro del cual se analiza el caso de Chile. Finalmente, la autora identifica los factores críticos relacionados con el uso de los resultados de la evaluación del rendimiento escolar.

A continuación se presenta un esquema conceptual para analizar los sistemas nacionales de evaluación del rendimiento escolar y el uso de la información que tales sistemas generan y aplicar dicho esquema a los sistemas nacionales de evaluación de la educación chilena. Específicamente se ejemplifican aquellos casos

en que puede rastrearse claramente la utilización de la información generada por los sistemas de evaluación. En dos de estos sistemas se efectúan mediciones en el ámbito de la educación primaria y, en uno de ellos, al término de la educación secundaria. El documento identifica, además, algunos factores que han favorecido o inhibido el uso de los resultados de la evaluación y propone algunos principios que pueden ayudar a orientar las políticas nacionales de evaluación.

SISTEMAS NACIONALES DE EVALUACIÓN Y USO DE LA INFORMACIÓN

En los últimos diez años se ha producido un amplio debate sobre la calidad de la educación en América Latina. Lo anterior es explicable puesto que, en las décadas precedentes, los países de la región centraron todos sus esfuerzos en la expansión de la cobertura de los sistemas educativos, lo que relegó el tema de la calidad a un segundo plano. En los países en los que el problema de la cobertura ha sido gradualmente resuelto, se ha detectado que la expansión de los sistemas educativos se realizó a costa de la calidad del servicio educativo, porque la inversión requerida superaba con creces las posibilidades de las economías de dichos países. Es en este contexto que ha surgido un renovado interés político sobre el problema de la calidad. Este interés surge a partir del reconocimiento de la importancia del capital humano en el desarrollo de las economías. A ello se suma el apoyo financiero decidido que los organismos internacionales (por ejemplo, el BID, la OEA y el Banco Mundial) han brindado al sector educación y, cuyo impacto sobre el mejoramiento de los sistemas educativos locales, quieren ver demostrado a través de resultados tangibles (Lockheed, 1991).

La evaluación del rendimiento estudiantil surge, entonces, como una necesidad cuando un sistema educativo logra la igualdad de oportunidades en el acceso a la educación y empieza a cuestionarse la calidad del servicio educativo que se ofrece en un país. En este contexto comienzan a ponerse en marcha los sistemas nacionales de evaluación que permiten suministrar información referente al logro de las metas educativas; identifican variables internas y externas al sistema que explican las desigualdades en los resultados; contribuyen a una predicción informada del funcionamiento del sistema en el futuro y, proporcionan indicadores acerca de los rasgos más permanentes del sistema.

Los sistemas de evaluación comprenden, en general, la aplicación de pruebas o tests referentes al logro de los objetivos de aprendizaje en las áreas que constituyen los ejes del currículo escolar, complementadas en ocasiones, con cuestionarios sobre variables potencialmente explicativas de las diferencias en los resultados información que se suelen diseminar entre los agentes educativos directos como los maestros y directores de escuela e, indirectos, como los medios de comunicación.

Entre los propósitos señalados más frecuentemente para estos sistemas de evaluación se citan los siguientes (Lockheed, 1996; Greany y Kellaghan, 1996):

- fundar y evaluar políticas educativas,
- evaluar programas educativos específicos,
- realizar el seguimiento de los cambios que experimentan los resultados educativos a través del tiempo,
- responsabilizar a educadores de los aprendizajes alcanzados por los estudiantes,
- seleccionar y distribuir a los alumnos hacia niveles educativos superiores,
- certificar los logros de aprendizaje de los estudiantes,
- proporcionar datos a los padres y apoderados sobre la calidad de la educación ofrecida por las escuelas y,
- diagnosticar necesidades de aprendizaje.

No obstante, difícilmente se pueden lograr todos estos propósitos a través de un sólo sistema de evaluación. Por tal motivo, a menudo coexisten dos o más sistemas con finalidades diferentes.

En todo caso, se espera que un sistema educativo que persigue algunos de los propósitos enunciados, contribuya a una mejor comprensión del funcionamiento del sistema educativo, proporcione las orientaciones necesarias para la toma de decisiones de diversos actores en diferentes niveles y, coadyuve al mejoramiento de la calidad del servicio educativo. En otros términos, se supone que la evaluación del rendimiento escolar tiene un impacto sobre el propio sistema educativo más allá de proporcionar información sobre el mismo. Se espera que las informaciones orienten una toma decisiones conducente a acciones, cuyos efectos a su vez pueden ser evaluados por los mismos sistemas, siempre que éstos operen a lo largo del tiempo. Dependiendo de cuáles sean los objetivos de un sistema de evaluación del rendimiento escolar, éstos pueden tener un impacto en diferentes áreas o sectores (Figura 1).

Figura 1. Algunas áreas sobre las que pueden influir los Sistemas Nacionales de Evaluación del Rendimiento Escolar.

De este modo, si un sistema de evaluación del rendimiento tiene los objetivos de generar informaciones para fundar y evaluar políticas educativas y llevar a cabo un seguimiento de sus resultados educativos, los resultados de la evaluación pueden influir en la propuesta de nuevas políticas e incluso reformas educativas. Por otro lado, a su vez es posible que tengan efectos sobre la gestión pedagógica de las escuelas y las decisiones que adopten los maestros sobre la conducción del proceso educativo.

Cabe señalar, sin embargo, que estos efectos tienen lugar en la medida en que efectivamente se utilicen los resultados de las evaluaciones para la toma de decisiones—lo que es un supuesto que no siempre se cumple. En la literatura especializada, diversos autores (Alkin, 1979, 1985; Brown, Newman y Rivers, 1985) reconocen dos formas de conceptualizar el uso las informaciones, a saber, la denominada perspectiva de la corriente principal y la concepción alternativa.

La primera postura concibe el uso como el impacto directo y rápido de la información evaluativa sobre el sistema educativo, programa o, en general, el objeto de la evaluación. Bajo esta concepción, el uso es considerado como un evento—no como un proceso— susceptible de ser iniciado desde el momento en que se planifica un sistema de evaluación. De esta manera se dicotomiza la dimensión del uso en dos categorías extremas: uso versus no uso. Adherirse a esta corriente implica presumir que se acepta la utilización de la evaluación, sólo cuando ésta produce efectos tales como la introducción de reformas educativas inmediatas, los cambios de un programa por otro o las modificaciones drásticas en las estrategias pedagógicas. Estas acciones radicales no ocurren necesariamente, ya que existe un conjunto de factores que condicionan el empleo de la información evaluativa, que dependen de las características del proceso evaluativo y de los resultados que éste proporciona. Asimismo, debe considerarse que en la toma de decisiones influyen otros factores, además de los resultados de las evaluaciones.

King y Peachman (1984), en un intento de clarificar esta concepción de uso, indican que en la base de ella se encuentran algunos supuestos cuestionables, que se describen a continuación:

- considerar que las decisiones se pueden adoptar de un modo clásicamente racional, sin tener en cuenta las variables políticas, sociales y organizativas que influyen en ellas;
- entender que la información evaluativa constituye el único factor desencadenante de efectos inmediatos y observables (el mito del big-bang),
- estimar que la calidad de los informes evaluativos es una condición suficiente para su uso cabal y,
- juzgar que la colaboración activa entre los evaluadores y responsables de las decisiones incrementará necesariamente el uso.

Al respecto, se puede mencionar que los gestores del Programa de Evaluación del Rendimiento (PER) que se desarrollara en Chile entre 1982 y 1984 comenzaron la

operación del programa precisamente con la adopción de la mayoría de estos supuestos, ya que supusieron erróneamente que, a partir de dicha información, los maestros, directores de escuela y autoridades educativas, generarían propuestas de acciones a todo nivel, mediante un proceso de autodirección y autocontrol. No obstante, en el corto plazo se dieron cuenta de la debilidad de los supuestos mencionados y se adhirieron a la concepción alternativa.

Otro ejemplo que ilustra esta postura se encuentra en Schiefelbein (1992: 264), quien juzga el impacto del PER sólo a partir de la constatación de que los resultados alcanzados por los alumnos en las evaluaciones no mostraron variaciones significativas, señalando lo siguiente: "estos antecedentes muestran que las expectativas, en cuanto a mejoría de los rendimientos académicos, son demasiado altas".

En cambio, el enfoque alternativo del uso de la información evaluativa lo concibe como un proceso gradual en el que dicha información, junto con otros antecedentes, puede ir generando pequeñas acciones, las cuales modifican paulatinamente la situación inicial detectada. Bajo esta concepción, el impacto de la evaluación no es inmediato, puede demorarse años y producirse cuando es combinado con otras informaciones contextuales, o bajo distintas circunstancias, lo que redunda en que pueda adoptar significados diversos en distintos tiempos (Braskamp, 1982).

En la opinión de King y Pechman (1984) existen al menos tres niveles de uso de la información evaluativa, a saber:

- simbólico o persuasivo
- conceptual
- instrumental

A estos niveles se puede agregar una categoría de uso no instrumental. La relación entre la producción de datos evaluativos, las informaciones contextuales y su empleo como insumo de la toma de decisiones se sintetiza en el diagrama contenido en la Figura 2.

Cuando el empleo de los resultados de las evaluaciones se ubica en el nivel simbólico o persuasivo, se utiliza con fines personales. Por ejemplo, los hallazgos de una evaluación pueden ser usados por un director de escuela para justificar su buen desempeño, con lo cual dispone de antecedentes para su continuidad en el cargo. Esta modalidad exige un análisis del contexto en que opera, para juzgar adecuadamente las intenciones positivas o negativas subyacentes. En este caso, la evaluación es utilizada para justificar ciertas decisiones más que como insumo de las mismas y, a veces, suele usarse en este sentido por razones político-administrativas.

A su vez puede incluirse en este primer nivel, un efecto de los sistemas de evaluación del rendimiento, ampliamente discutido en la literatura (Greany y

Kellaghan, 1996), esto es, la forma en que influyen sobre la enseñanza. De hecho, ninguna aplicación de pruebas externa es neutral y los maestros, cuando de algún modo perciben la importancia de los resultados de las pruebas, comienzan a orientar el proceso de enseñanza predominantemente hacia los contenidos y objetivos incluidos en las pruebas y comienzan a justificar su forma de abordar el currículo escolar, a partir de la cobertura de las pruebas.

Figura 2. Usos de la información evaluativa

En cambio, el uso conceptual implica que la evaluación provoque en el usuario una reflexión más detenida acerca del objeto de evaluación, de forma que pueda reconocer que existen aciertos o dificultades para su desarrollo, conduciendo de esta

manera a una modificación de sus actitudes. El proceso de cambio actitudinal puede llevarlo a largo plazo a tomar algunas decisiones muy específicas. Por ejemplo, el director de una escuela puede atribuir los bajos resultados en una evaluación del rendimiento a una falta de identificación de los alumnos con la escuela. Ante esta conclusión, convoca a los profesores a una jornada de discusión para buscar las causas de esta dificultad. Este ejemplo muestra, que si bien no se han emprendido acciones específicas, la información ha incentivado al usuario a una reflexión sobre el tema, la que después, eventualmente, se traduce en pequeñas acciones como encuentros informales entre maestros y alumnos.

Por otra parte, se califica el uso como instrumental, cuando se pueden vincular claramente las informaciones generadas por la evaluación con decisiones tomadas por un usuario directo de las mismas. Empleando el mismo ejemplo anterior, si el director decide recabar sistemáticamente las opiniones de los alumnos de la escuela acerca de los aspectos que consideran positivos y negativos en la misma para luego poner en marcha un programa conducente a mejorar la identificación de los estudiantes con el plantel, entonces estaría utilizando instrumentalmente la información proporcionada por el sistema de evaluación.

Por último, si la información ha sido desechada conscientemente por los usuarios, se puede denominar como «no uso instrumental». Si por ejemplo, el informe de resultados de la evaluación del rendimiento muestra que los alumnos de un municipio han obtenido un desempeño extremadamente bajo en ortografía en el cuarto grado y el departamento técnico pedagógico de dicha instancia estima que la ortografía literal no es un objetivo educativo fundamental en ese grado, entonces, probablemente no se efectuará ninguna acción para mejorar el logro de los aprendizajes del objetivo mencionado.

Al analizar el uso de la información de acuerdo a este modelo, es posible reconocer la presencia de los efectos de un sistema nacional de evaluación del rendimiento escolar en múltiples acciones, las que no necesariamente implican un mejoramiento rápido de los rendimientos. No obstante, dichas intervenciones pueden llevar a mediano o largo plazos a elevar la calidad de los aprendizajes estudiantiles.

En este estudio, se denominará impacto social a las formas de uso de los resultados de las evaluaciones que influyen en decisiones conducentes a acciones, que tienen efectos o consecuencias para grupos de la sociedad, ya sea dentro o fuera del sistema educativo. A lo que se puede agregar que, de hecho, la influencia de la evaluación es difícilmente observable en la toma de decisiones propiamente tal, mas se puede inferir a partir de algunas señales explícitas en la acción que sigue.

Descripción de los sistemas de evaluación del rendimiento escolar en Chile

En esta sección se describen los tres sistemas de evaluación de mayor permanencia en el tiempo en Chile. El primero de ellos, Programa de Evaluación del Rendimiento Escolar (PER), se centra en la educación básica o primaria. Se incluye

en esta discusión porque sirvió de base para el desarrollo del segundo, el Sistema Nacional de la Medición de Calidad de la Educación (SIMCE), que se ha venido aplicando desde 1988. El SIMCE incorporó por una vez la medición en el segundo grado de la educación secundaria. El tercero es el Sistema Nacional de Ingreso a la Educación Universitaria, que se aplica al término de la educación secundaria o media y funciona desde 1967.

Programa de Evaluación del Rendimiento Escolar (PER)

La idea de medir la calidad de la educación nació en Chile en 1978 en el Ministerio de Educación (MINEDUC), el cual solicitó a la Pontificia Universidad Católica que diseñara y pusiera en marcha un sistema de información educativa. En el marco de la descentralización administrativa que comenzaba a ejecutarse, este programa permitiría recoger información sobre la calidad de la educación y suministrarla a los diversos agentes educativos para que pusieran en marcha acciones destinadas a su mejoramiento.

Esta tarea fue encomendada, por medio de un convenio, a un equipo interdisciplinario, que se hizo cargo, tanto del estudio de viabilidad como el de la puesta en marcha del PER. Este equipo comprendía un pequeño grupo de profesionales (la mayoría de ellos con estudios de postgrado en los E.E.U.U.), profesores de la Universidad Católica. Entre ellos, cuatro ingenieros, tres psicólogos, tres educadores y un sociólogo. Los psicólogos y educadores tenían la responsabilidad de la elaboración de los diferentes instrumentos de medición educativa, al tiempo que llevar a cabo la coordinación con las autoridades educativas y con la red administrativa-educacional en las regiones. Por su parte, los ingenieros tenían a su cargo la administración, logística y computación del PER (Himmel, 1996).

El objetivo general del PER fue el de contribuir a la mejora de la calidad educativa a través de la descentralización de las responsabilidades. La expectativa del equipo era que, como consecuencia de la evaluación del rendimiento, todos los agentes del proceso educativo asumirían un papel más activo en el mismo. El supuesto subyacente fue que el sólo hecho de proporcionar información acerca del logro de los objetivos fundamentales de los alumnos induciría a los profesores y directores a generar diversas innovaciones para mejorar la calidad de la educación.

Por otro lado, la disponibilidad de información objetiva, confiable y válida para las autoridades educativas centrales permitiría enfocar las políticas educativas de un modo más realista.

Los objetivos específicos formulados para el PER fueron los siguientes:

- Formular explícitamente los objeticos del aprendizaje, cuyo logro fue considerado como fundamental para la educación primaria por el MINEDUC.
- Proporcionar información acerca del logro de dichos objetivos educativos a

padres, profesores, directores y autoridades a niveles municipal, regional y central, con el nivel de especificidad y agregación apropiado a las necesidades de cada audiencia.

- Suministrar al Ministerio la información necesaria para desempeñar sus nuevas funciones normativa y de supervisión. Dicha información estaría destinada a identificar aquellas escuelas que, por sus bajos resultados, requerían de mayor apoyo técnico-pedagógico y financiero. Además, permitiría efectuar un seguimiento de las escuelas y municipios de acuerdo al logro alcanzado y, asimismo, redefinir el contenido de la planificación central, en lo que concierne al desarrollo del currículo, perfeccionamiento de maestros y diseño de textos de estudio y material de enseñanza.

El diseño del Programa tuvo los siguientes rasgos principales (Himmel, 1996):

- El logro de los objetivos educativos fue medido en cuarto y octavo grados. Se seleccionó el cuarto curso porque constituye el término del primer ciclo de la Educación General Básica y el último en que todas las asignaturas son impartidas por un sólo profesor. A su vez, el octavo curso marca el término de la escolarización obligatoria en Chile y, en este segundo ciclo, las asignaturas son impartidas por diferentes maestros.
- Dado que se consideraron como pilares del proceso educativo primario las habilidades y destrezas que logran los estudiantes en lenguaje (comprensión lectora y expresión escrita) y matemáticas (adquisición de algoritmos y resolución de problemas), se aplicaron versiones completas de las pruebas respectivas a cada estudiante. El resto de las mediciones—en ciencias naturales, sociales y el dominio socioafectivo—se efectuó de forma matricial, esto es, cada alumno respondía a una muestra de preguntas, pero el promedio de la clase daba cuenta del aprendizaje del área respectiva.
- Todas las pruebas, con excepción de la de expresión escrita, fueron progresivas y de selección múltiple—esto es, cubrían los objetivos educativos del ciclo completo.
- La construcción de las pruebas se efectuó conforme a rigurosos estándares internacionales por profesores de aula, como por especialistas en currículo y evaluación. Las pruebas se centraban en una selección de objetivos fundamentales y contenidos mínimos. Las versiones definitivas de todas las pruebas de selección múltiple tuvieron una confiabilidad superior a 0,90 en las pruebas del dominio cognoscitivo y superior a 0,80 en las correspondientes al desarrollo socioafectivo. La prueba de redacción tuvo una confiabilidad intercorrectora superior a 0,80 con el método analítico.
- Se optó por aplicar pruebas prácticamente censales, por cuanto se postuló que los agentes educativos sólo emprenderían acciones frente a los resultados referentes a su propio medio, efecto que sería mucho menor con respecto a los resultados de una muestra. Sólo quedaron fuera de las aplicaciones escuelas muy pequeñas (con menos de 5 alumnos en los cursos respectivos) y las que se encontraban en lugares de difícil acceso durante la mayor parte el año. Las

pruebas se aplicaron aproximadamente a un total de 400.000 alumnos por año, lo que representaba una cobertura del 90 por ciento de los estudiantes.

- Conforme a los criterios que se desprendían de investigaciones previas (Himmel et al., 1980; Himmel et al., 1982), las escuelas fueron segmentadas en grupos homogéneos de acuerdo al nivel socioeconómico de la población que atendían. El fin era separar el efecto de las acciones que emprendieran las escuelas de la influencia de las variables estructurales. Estos grupos se denominaron "estructuras socioeconómicas".

  Se hizo un extenso análisis sobre la ventaja que podría significar el facilitar información sobre el promedio de los indicadores socioeconómicos de los alumnos por escuela junto con los resultados de las pruebas—que era el procedimiento adoptado en los E.E.U.U. (Alkin, 1981). No obstante, se llegó a la conclusión de que ésta no constituía una solución mejor que la de las estructuras, pues su interpretación resultaba más compleja para los usuarios.

  La definición de las estructuras socioeconómicas fue considerada como una decisión administrativa importante para proporcionar un referente comprensible, que permitiera a las audiencias, en particular los directores y maestros, una interpretación y análisis más directos.
- La difusión previa a la aplicación de las pruebas consistió en la distribución de trípticos explicativos de los objetivos del PER a directores, profesores y padres. Además, se distribuyó cada año un folleto técnico que describía y ejemplificaba los instrumentos de medición que se aplicarían y se preparó material audiovisual que fue difundido por el personal de la red administrativa a las autoridades locales, los directores y profesores.
- Las pruebas se administraban al término del año escolar, bajo condiciones controladas para garantizar la homogeneidad del proceso. Para este efecto se organizó una red administrativa de supervisión de 640 personas que abarcaba todo el territorio nacional y se preparó a 12.000 examinadores. Además se tomaron medidas muy estrictas para la seguridad de las pruebas.
- Los resultados de las aplicaciones de pruebas se hacían públicos durante el primer mes de clases del año inmediatamente siguiente. La información consistía en el porcentaje promedio de respuestas correctas en cada prueba y el mismo indicador para cada conjunto de objetivos. Asimismo, se prepararon informes con diversos niveles de agregación para las diferentes audiencias. A los maestros, por objetivos y totales por clase; a los directores, por escuela y grado y a las demás autoridades según el nivel en que se desempeñaban, bien fuera municipal, regional o nacional

  En los informes se incluían, además de los resultados del nivel correspondiente, dos indicadores para la interpretación comparativa. El primero consistía en el percentil 25 y 75 de los resultados alcanzados por los alumnos de las escuelas en la misma estructura y, el segundo, en el percentil 5 y 95 a nivel nacional. En las mediciones subsiguientes se proporcionó información comparativa con los años precedentes.
- Se prepararon y distribuyeron manuales de interpretación de los resultados de las pruebas y recursos audiovisuales a las autoridades nacionales, locales y a los directores y maestros.

- Se suministraron informes específicos al MINEDUC acerca de las escuelas que requerían más urgentemente apoyo técnico y económico.
- Como complemento a toda esta información, se prepararon manuales de orientación pedagógica con respecto a los objetivos que, a nivel nacional, mostraban el más bajo logro. Estos manuales se distribuyeron a todos los profesores de los grados evaluados.
- Se propuso al Ministerio un conjunto de recomendaciones compatibles con el diseño del PER para que, a través de la supervisión escolar, se brindara el apoyo técnico-pedagógico a las escuelas de más bajo rendimiento.
- El equipo del PER introdujo un conjunto de innovaciones en la medición educativa, que siempre fueron empíricamente contrastadas, sobre todo cuando iban en contra de creencias profundamente arraigadas en el sistema educativo. Así, por ejemplo, en una encuesta a profesores se encontró que éstos sostenían que los alumnos de cuarto curso no podían responder pruebas en hojas de respuesta. Se experimentó la hoja de respuesta en una muestra de alumnos de muy bajo nivel socioeconómico y se encontró que la diferencia entre los promedios de los niños que contestaban en hoja de respuestas y los que respondían directamente en el folleto, no era significativa.

El convenio entre el Ministerio de Educación y la Pontificia Universidad Católica de Chile había sido suscrito inicialmente por un período de un año para efectuar el estudio de viabilidad de la puesta en marcha del PER y, luego, se renovó por tres años para su puesta en marcha. Terminado este período, el MINEDUC no renovó más el convenio, con lo cual se suspendió el PER.

Aunque en el breve período de tres años no fue posible detectar una mejora substancial en el logro de los objetivos educativos, se observó que, para algunos de los objetivos en los que el logro en la primera medición fue muy bajo, se produjeron incrementos importantes como consecuencia de las acciones remediales emprendidas por profesores y escuelas. Tal es, por ejemplo, el caso de la resolución de problemas verbales en matemáticas, el que en ambos grados medidos, experimentó un aumento de casi un 10 por ciento.

Cabe señalar que, durante el período de implantación del PER, se sucedieron seis Ministros de Educación—con los consiguientes cambios de autoridades superiores del Ministerio. El último de éstos anunció a comienzos de 1984, año de culminación del convenio con la Pontificia Universidad Católica, que el PER dejaría de aplicarse. Nunca discutió esta decisión con el equipo encargado del proyecto y su fundamentación permanece hasta hoy en el plano especulativo, pues todos los intentos de negociación por parte de las propias autoridades del Ministerio, que apoyaban el Programa, no prosperaron en hacer cambiar esta decisión.

Se puede sostener que uno de los factores desencadenantes de la terminación del programa fue su costo que ascendía a 5 dólares americanos por alumno, esto es, unos dos millones de dólares, que en un período de recesión pueden haber

impresionado a las autoridades. Un estudio de seguimiento del PER (Himmel et al., 1988), puso de manifiesto que las más altas autoridades del período no estaban comprometidas con el PER y consideraban que el programa tenía un costo demasiado elevado.

A modo de conjetura, es probable que existiera una tensión entre posturas antagónicas en el interior de la burocracia ministerial, una a favor y la otra en contra del manejo del PER por parte de la Universidad Católica, pues en el interior del MINEDUC había más de un grupo que se sentía perfectamente capaz de ejecutar el proyecto.

Otra explicación plausible se refiere al contexto político más amplio, dominado en ese entonces por un conjunto de economistas que promovían la economía neoliberal, mientras una alta proporción del profesorado se adhería al principio del "Estado docente", es decir, de radicar la función educativa en el Estado, que era una postura de larga tradición en Chile. A todo ello cabe añadir que, en ese momento, el proceso de descentralización administrativa estaba prácticamente paralizado.

Al término del convenio, se efectuó un estudio de seguimiento de las acciones que habían tenido lugar durante la aplicación del PER (Himmel et. al. 1988). Dicha investigación permitió constatar que en algunas de las escuelas se habían emprendido acciones concretas tendentes a mejorar el aprendizaje de los estudiantes. Entre ellas se pueden mencionar: el análisis exhaustivo de los programas de curso para establecer jerarquías de aprendizaje y la revisión de las formas de evaluación utilizadas por los profesores. Esto generó un cambio en la «cultura evaluativa» y la puesta en marcha de algunas medidas administrativas, como el aumento en el número de horas asignado a lenguaje y matemáticas (Contreras, 1988).

Descripción del Sistema Nacional de Medición de la Calidad de la Educación (SIMCE)

En 1988 se completó el proceso de descentralización del sistema educativo. No obstante, en el MINEDUC todavía existía un sector importante que estimaba que tenía la capacidad técnica para desarrollar un sistema de evaluación del rendimiento estudiantil si se le proporcionaban los recursos económicos necesarios. Las autoridades consideraron que, para volver a poner en marcha un sistema de evaluación, no se requería meramente de recursos financieros adicionales, sino que, básicamente, era necesaria una capacidad técnica que no existía en el Ministerio.

Con el propósito de lograr la aceptación política del sistema de evaluación y la reducción de su costo, las autoridades superiores del Ministerio propusieron las siguientes medidas para reimplantarlo y lograr su aceptación (Himmel, 1996):

- Cambiar el nombre del sistema de evaluación para evitar su asociación con el PER. En adelante se denominaría Sistema de Medición de la Calidad de la

Educación (SIMCE).

- Suscribir un nuevo convenio con la Pontificia Universidad Católica, encargándola de dirigir por tres años la mayor parte de la ejecución del proyecto, pero con el compromiso de capacitar al equipo del MINEDUC, que se haría cargo del SIMCE una vez finalizado el convenio. Durante el cuarto año de implantación del proyecto, la ejecución sería asumida totalmente por el equipo del Ministerio y asesorada por el equipo profesional de la Universidad. Al término del convenio, el SIMCE quedaría institucionalizado en el MINEDUC.
- Centrar las mediciones censales, básicamente, en lenguaje y matemáticas, y encargar su ejecución técnica a la Universidad, con un modelo análogo al del PER.
- Efectuar una medición muestral del 10 por ciento de la población estudiantil en Ciencias Naturales e Historia y Geografía. La elaboración de los instrumentos respectivos sería la responsabilidad del equipo universitario.
- La medición del desarrollo afectivo, así como la incorporación de mediciones sobre el grado de aceptación de la labor educativa por parte de alumnos, profesores y padres, y algunos indicadores de eficiencia escolar tales como repitencia, promoción y abandono, estaría a cargo de un equipo autónomo en el Ministerio.
- Aplicar las mediciones en los cursos cuarto y octavo en años alternados.
- Encargar al Ministerio el proceso de aplicación de pruebas. Esto es, la organización de la red administrativa y contratación de examinadores; la logística y computación serían diseñadas y estarían a cargo de la universidad.
- La sensibilización, difusión del sistema y diseminación de los resultados, se efectuaría con apoyo del equipo de la universidad.
- Definir como objetivos del SIMCE los siguientes:

  - Proporcionar asistencia al Ministerio de Educación en su tarea normativa y en la supervisión del sistema.
  - Brindar apoyo a las autoridades regionales y locales en la supervisión, además de prestarles apoyo técnico.
  - Estimar la calidad de cada establecimiento educativo; comparar dichos resultados, buscar los factores explicativos y evaluar los resultados de los programas pedagógicos.
  - Orientar la actividad de perfeccionamiento docente, las acciones de supervisión y la asignación de recursos.

Como puede advertirse, existe bastante similitud entre los propósitos del PER y los del SIMCE.

Desde entonces, ha funcionado regularmente el sistema, introduciéndose modificaciones en las áreas no propiamente académicas, como por ejemplo la medición de la creatividad y el desarrollo personal (Prado, 1995).

Entre las características actuales del sistema, que posiblemente sean peculiares de

éste, se puede destacar que desde 1995 se publican en los periódicos los resultados nacionales, regionales, locales y por escuela, lo que ha contribuido a una mayor difusión pública del SIMCE.

A lo expuesto, cabe agregar que el SIMCE, ya se encuentra institucionalizado, formando parte de la labor habitual de las escuelas. Los directores y profesores reconocen que el sistema tiene características técnicas apropiadas y es un medio útil para la gestión pedagógica, aunque señalan que no proporciona un cuadro completo de la calidad de las escuelas (Zabalza et al., 1994).

Descripción del sistema nacional de ingreso a la educación universitaria

Hasta 1966 el requisito legal de ingreso a las ocho universidades chilenas era el Bachillerato, que consistía en una versión nacional del *Baccalaureat* francés y era administrado por la Universidad de Chile desde 1927. En ese año, el Congreso Nacional suprimió este requisito y las universidades se vieron obligadas a diseñar un sistema de selección de acceso.

La Universidad de Chile había desarrollado y validado un proyecto de sistema de selección desde 1962, que incluía un conjunto de pruebas objetivas de selección múltiple y tenía en cuenta las calificaciones alcanzadas por los candidatos en la educación secundaria (Lara, 1985).

Suprimido el Bachillerato y, ante la necesidad de acometer la selección de estudiantes para la universidad en un período muy breve, se conformó una comisión interuniversitaria ad hoc, con amplias atribuciones para resolver el problema. Dicha comisión delegó a la Universidad de Chile la administración del Sistema de Selección, ya que ésta contaba con la experiencia técnica para llevarlo a cabo. La comisión informal se transformó luego en la Comisión Coordinadora del Proceso de Admisión del Consejo de Rectores de las Universidades Chilenas, que subsiste hasta el presente y tiene como principal función velar por el funcionamiento correcto del proceso de selección.

Este sistema de ingreso ha sufrido pocas variaciones en sus 30 años de funcionamiento y sus rasgos más importantes son los que siguen:

- Hasta 1982, las pruebas del sistema de selección eran un requisito de ingreso para las ocho universidades que existían en el país. Actualmente, sólo son un requisito para las universidades que reciben aportes financieros del Estado, que son las ocho tradicionales más sus derivadas. Muchas de las universidades privadas sin financiamiento estatal no las exigen.
- Está compuesto por un conjunto de pruebas de selección múltiple obligatorias y otro de la misma naturaleza de carácter electivo entre las que el postulante a la universidad puede seleccionar aquéllas que son requisito para el programa de estudios que intenta proseguir. No obstante, cualquier candidato puede presentarse a todas las pruebas, pues su aplicación está programada de modo que

no se superponga. Una de las pruebas obligatorias es la Prueba de Aptitud Académica (PAA), que es una prueba similar al SAT norteamericano y que se compone de una parte verbal y una matemática.

- La otra prueba, a la que necesariamente deben someterse en la actualidad todos los candidatos, es la Prueba de Historia de Chile, que fue incorporada al sistema en 1984 para fomentar el aprendizaje de la misma durante la secundaria.
- Las pruebas electivas, designadas como Pruebas Específicas de Conocimientos, responden a las necesidades planteadas por los programas universitarios y consisten en pruebas que cubren cinco de las asignaturas de la educación secundaria, a saber: matemáticas, biología, física, química y ciencias sociales. Durante una época se aplicó solamente una prueba de ciencias, que comprendía biología, física y química.
- Todas las pruebas son expresadas en una escala estándar de promedio 500 y desviación estándar 100.
- El otro antecedente que se incorpora como criterio de selección son las calificaciones promedio de la educación secundaria que también se expresan en la misma escala que las pruebas.
- Las pruebas se aplican una sola vez al año, al final del período escolar. El estudiante se inscribe en las pruebas unos tres meses antes de rendirlas y los resultados se publican con el número de la cédula de identidad de los postulantes, en un periódico de circulación nacional, un mes y medio después de haberse rendido.
- Una vez conocidos los resultados, cada candidato postula, mediante un sólo trámite, a cualesquiera de las universidades tradicionales o sus derivadas. El candidato puede postular a varios programas de diferentes universidades.
- La selección propiamente tal es realizada por la Universidad de Chile, de acuerdo a los criterios y ponderaciones establecidos por cada universidad. Los criterios y ponderaciones son establecidos por la mayoría de las universidades a partir de los estudios de validez predictiva que realizan y sólo tienen como restricción ponderar la prueba de historia de Chile en el 10 por ciento y las calificaciones de la educación secundaria en el 20 por ciento. El resultado de esta selección se publica, a su vez, en un periódico de circulación nacional, por universidad y, dentro de cada una de ellas, por programa.
- La coordinación total del proceso se encuentra a cargo de la comisión ya mencionada.

El sistema se ha aplicado por 30 años y, a pesar de que periódicamente surgen voces que reclaman el cambio radical de éste, hasta ahora no se han presentado propuestas que compitan en eficiencia y costo con la opción vigente.

Impacto social de los sistemas de evaluación del rendimiento escolar en Chile

Impacto social del SIMCE en las políticas educativas

A continuación se describen tres líneas de la política educativa chilena que fueron fundadas con base en los resultados del SIMCE; éstas son: el Programa de las 900 escuelas (P-900); el programa de mejoramiento de las escuelas rurales (MECE-RURAL) y la línea de Proyectos de Mejoramiento Educativo (PME). Estas iniciativas se inscriben en el uso instrumental de las informaciones proporcionadas por el SIMCE.

El programa de las 900 escuelas

Este programa se encuentra focalizado hacia las escuelas de los sectores pobres y fue el primero que se puso en marcha en 1990, al iniciar su mandato el gobierno democrático. Se orientó hacia el 10 por ciento de las escuelas que presentaban los más bajos logros en el SIMCE. Estas escuelas atienden a una población escolar en extrema pobreza. Este diagnóstico se efectuó a partir de los resultados del SIMCE de 1988. Se postuló como una discriminación positiva enfatizando los aprendizajes en las áreas de lenguaje y matemáticas de los alumnos de primero a cuarto grados. Tiene asimismo, el propósito de promover la creatividad y el desarrollo personal y social de los educandos. A la vez, se capacita a los maestros en metodologías que faciliten tanto la integración de los alumnos en su medio cultural, como el desarrollo de la cooperación entre la escuela y la comunidad.

El programa se encuentra estructurado con los siguientes componentes :

- Los Talleres de Perfeccionamiento Docente, de carácter participativo, orientado a la actualización teórica, reflexión sobre la práctica con vistas a introducir innovaciones en el aula y adquisición de una postura profesional frente a los resultados obtenidos por los alumnos.
- Los Talleres de Aprendizaje (TAPs) brindan atención especial a los alumnos de tercero y cuarto grados que presentan retraso pedagógico. Los TAPs están dirigidos por jóvenes monitores entre 18 y 25 años de la comunidad en que se encuentra la escuela, los cuales reciben capacitación permanentemente en las áreas en las que ellos detectan problemas en el trabajo con los niños (Cabezón, Condemarín y Vaccaro, 1996). Los TAPs se desarrollan en la jornada alterna a la escolar.
- La dotación de materiales educativos, que comprende la distribución de bibliotecas de aula y materiales didácticos, especialmente juegos, para apoyar el aprendizaje en lenguaje y matemáticas, en las escuelas beneficiadas por el programa. Entre los materiales didácticos también se ha distribuido un gran número de calculadoras de bolsillo.
- El mejoramiento de la infraestructura, que consiste en la reparación y mejora del local escolar en los aspectos considerados por los maestros como el mejor apoyo al aprendizaje escolar.
- La supervisión técnico-pedagógica. Este componente se encuentra a cargo de los supervisores del MINEDUC, los cuales han recibido, a su vez, una capacitación para el desempeño de sus funciones en el P-900.

Las escuelas se incorporan al programa en la medida que sus resultados en el SIMCE se encuentren muy por debajo del promedio regional y egresan del programa cuando los resultados evaluativos se encuentran sobre dicho promedio, o cuando han incrementado sus logros en mayor medida que el resto de las escuelas gratuitas. Este hecho revela que los resultados en el SIMCE no sólo sirven para que las escuelas accedan al P-900 sino, también, como un criterio de impacto del programa. El número de escuelas beneficiadas con el programa ha pasado de 900 en 1990, a aproximadamente 1.300 en 1991. Más adelante este número disminuyó a unas 1.000 en 1996. La fluctuación en el número de las escuelas es un indicador bruto del mejoramiento experimentado por un grupo de escuelas, pues implica que han logrado mejores resultados en el SIMCE.

Cabe mencionar, además, que en el marco del programa se han capacitado casi 14.000 jóvenes como monitores comunitarios. Estos jóvenes pertenecen a los mismos sectores pobres de las escuelas y han logrado, a través de la capacitación, incrementar sus competencias básicas y habilidades sociales.

Programa de mejoramiento de las escuelas rurales (MECE-RURAL)

Este programa es componente de una de las líneas de acción del Programa de Mejoramiento de la Calidad y Equidad de la Educación (MECE). Está encaminado hacia el mejoramiento de la calidad del servicio educativo ofrecido en las escuelas de alto riesgo educativo ubicadas en el sector rural. Este sector constituye el 20 por ciento de la población de Chile y presenta una gran heterogeneidad derivada de la geografía del país. Una alta proporción de la población rural se concentra en dos regiones del sur de Chile, cuyos resultados en las evaluaciones nacionales han sido extremadamente bajos. Dentro de la línea de acción del P-900, se llevó a cabo en 1992 una experiencia piloto que dio origen a la puesta en marcha del MECE-RURAL en 1993, que se centra en las escuelas uni, bi y tridocentes. (Parraguez, 1993).

Los principales aspectos que incluye este programa son:

- Diseño de una propuesta metodológica, a través de la formulación de un Manual de Desarrollo Curricular, cuyo propósito es articular el conocimiento general con la cultura local con el fin de hacer más pertinente el currículo escolar en cada región.
- Capacitación de los docentes para poner en marcha innovaciones pedagógicas orientadas a mejorar el aprendizaje de los niños de las zonas rurales y, asimismo en las estrategias pedagógicas congruentes con dichas propuestas, por ejemplo, el uso de textos escolares y de material didáctico.
- Elaboración y dotación de textos de primero y sexto grados, con un diseño apropiado para el medio rural y acorde con las nuevas propuestas metodológicas.
- Elaboración y dotación de material didáctico para profesores y alumnos, que contribuya a enriquecer el aprendizaje en las áreas de lenguaje, matemáticas y ciencias.

- Construcción de salas y contratación de profesores para completar las escuelas cuya cobertura no alcanza al octavo grado, que es el último de la escolarización obligatoria.
- Formación de "microcentros rurales" de coordinación pedagógica, que son agrupaciones de todos los maestros de un conjunto de escuelas geográficamente próximas. Tienen como propósito brindar a los maestros un espacio para que puedan reunirse periódicamente a reflexionar sobre sus prácticas pedagógicas, intercambiar y formular nuevas experiencias y evaluar su desarrollo, así como para recibir apoyo técnico pedagógico de los supervisores.
- Al igual que el P-900, este programa incorpora un fuerte componente de supervisión técnico-pedagógica y comprende una capacitación de los supervisores.

El porcentaje de escuelas uni, bi y tridocentes incorporadas al SIMCE era relativamente bajo hasta 1996, año en el que se incluyó una proporción mucho mayor de ellas. Al respecto, un informe publicado por el MINEDUC destaca que, entre 1994 y 1996, los resultados de cuarto grado en las escuelas incorporadas al MECE-RURAL subieron seis puntos en lenguaje y ocho en matemáticas (Ministerio de Educación, 1997), aunque no se trató de una medición dentro de un esquema experimental con grupos de control. Las pruebas que se utilizan en el MECE-RURAL son las mismas que para todo el país.

Los resultados del SIMCE se están utilizando para la identificación de problemas y, posteriormente, para el seguimiento de los programas diseñados para superarlos.

Línea de Proyectos de Mejoramiento Educativo (PME)

Esta línea fue concebida por el MECE como una estrategia general de descentralización pedagógica en Chile, que es un país donde el sector educación mantuvo una tradición de alto grado de centralismo. De este modo, constituye una forma de profundizar la descentralización del sistema educacional, la que en sus primeras etapas sólo estuvo orientada hacia la descentralización de la administración económica de las escuelas, dejando de lado la gestión pedagógica.

Los objetivos de dicha estrategia se encuentran centrados en el logro de un mayor aprendizaje por parte de los alumnos de la educación primaria que pertenecen a sectores pobres, a fin de disminuir la repitencia y la deserción escolar. Fue planteada de modo que cada escuela de educación general básica o primaria diseñara sus propios proyectos de innovación, orientados preferentemente a las áreas de lenguaje, matemáticas, ciencias naturales y ciencias sociales, y elaborados por los profesores de esas escuelas de forma participativa y autónoma.

La adjudicación de los proyectos se lleva a cabo mediante un concurso público al que tienen acceso prácticamente todas las escuelas básicas de administración municipal y las privadas gratuitas[1] subvencionadas por el Estado. Sólo tienen restricciones las escuelas que ya han realizado PME, se han incorporado

recientemente al programa P-900 o al MECE-RURAL. Las escuelas rurales uni, bi o tridocentes sólo pueden concursar a través de los "microcentros rurales".

La línea de los PMEs comprende la capacitación de los profesores para el diseño de los PMEs, la elaboración y ejecución de los proyectos, el perfeccionamiento en los contenidos y métodos de enseñanza durante la ejecución, la utilización de nuevos recursos de aprendizaje y la evaluación del proyecto.

Los supervisores técnico-pedagógicos, previamente capacitados en el programa PME, proporcionan asesoría a las escuelas para la elaboración de sus proyectos PME. Esta asesoría consiste en una capacitación a través de «talleres comunales» dirigidos hacia las escuelas territorialmente próximas y enfatiza los procedimientos de elaboración de proyectos—sin abordar los contenidos, a fin de estimular a las escuelas para que seleccionen los problemas que les son propios y propongan soluciones pertinentes.

Los proyectos generados por las escuelas son sometidos a una evaluación que se realiza en dos etapas de acuerdo a criterios claramente definidos. En la primera etapa, a nivel local, los PME son evaluados por un comité de supervisores capacitados por el Equipo de Gestión del MECE y en la segunda, a nivel central son ratificados y adjudicados a las escuelas.

Para la evaluación de los proyectos se emplean los criterios de calidad, equidad y pertinencia. La calidad se juzga en relación al potencial del PME para generar una mejora del proceso de enseñanza y aprendizaje que redunde en un incremento de los aprendizajes de los alumnos. El criterio de equidad privilegia a las escuelas que presentan mayores déficit de aprendizaje, medido básicamente a través del promedio de los resultados obtenidos por los alumnos en las pruebas de lenguaje y matemáticas del SIMCE. La pertinencia se refiere a la adecuación de los PMEs al ambiente socioeconómico y cultural en el que se inserta la escuela.

Conviene destacar también que, durante la etapa de diseño de los proyectos en las escuelas, uno de los antecedentes cruciales para la selección del problema es el análisis de los resultados del SIMCE de acuerdo al nivel de logro de los alumnos en los objetivos de aprendizaje.

A lo anterior se puede agregar que uno de los criterios de evaluación de la eficacia de esta estrategia es el de los cambios que se producen en las mediciones del SIMCE.

Como puede apreciarse, se ha utilizado una estrategia que genera este mejoramiento desde la base del sistema educativo hacia la cúspide de mismo. Se advierte además, que los resultados del SIMCE son utilizados como parte de los criterios de adjudicación de los proyectos y evaluación del éxito de los mismos, a la vez que sirven como uno de los insumos a los maestros para seleccionar las áreas de innovación en las que los logros de aprendizaje de los educandos son más débiles.

Esta última instancia de uso es un ejemplo de la forma en que se puede estimular el impacto de los sistemas de evaluación en los niveles del sistema, que se encuentran directamente involucrados en la provisión del servicio educativo.

Impacto social del SIMCE en los maestros y escuelas

La investigación acerca del impacto social del SIMCE en maestros y escuelas es relativamente escasa, pero fue posible recoger los resultados de dos de ellas, en las que se abordan este aspecto (Zabalza et al., 1994; Cerda et al., 1996).

En primer término, se detectan variaciones entre las escuelas rurales y urbanas con respecto a la gestión administrativa en las acciones derivadas de los resultados del SIMCE y que son explicables por el tamaño de las escuelas y las diferencias organizativas. Así, en las escuelas rurales, las intervenciones de mejoramiento son asumidas por el director de escuela junto a los maestros. En cambio, en el sector urbano son gestionadas por una estructura intermedia en la organización escolar, la Unidad Técnico-Pedagógica (UTP), con muy baja o nula participación del director. En el 29 por ciento de los establecimientos urbanos, el equipo directivo reconoce que no realiza ningún tipo de acción a partir de los resultados del SIMCE, ya que ello compete a la UTP (Zabalza et al., 1996). Este hecho puede actuar en contra del impacto de nuevas iniciativas, porque el liderazgo del director es identificado como un factor determinante en el éxito de las mismas (Rutherford, 1985; Davis y Thomas, 1989; Enrione, 1991).

Entre las acciones pedagógicas realizadas por las escuelas de ambos sectores, se encuentra la aplicación periódica de pruebas similares a las empleadas en el SIMCE, a fin de detectar el nivel de logro de los objetivos. También se formulan estrategias remediales y se lleva a cabo un seguimiento de estas iniciativas. Se agrega a esto, la institucionalización de tiempo complementario destinado principalmente a las áreas de lenguaje y matemáticas y la promoción del uso de nuevas metodologías (Zabalza et al., 1994).

En las escuelas rurales también es destacable el énfasis otorgado al desarrollo personal de los educandos, en particular rasgos como la autonomía, responsabilidad y autoestima. Los directores manifiestan que estos alumnos presentan debilidades en esta área, la que se evidencia cuando deben competir con los alumnos urbanos para ingresar en la educación secundaria (Zabalza et al., 1994).

Existen algunas escuelas en las que si bien se analizan los resultados del SIMCE, éstos no se usan para emprender innovaciones, principalmente con la excusa de la falta de recursos. En estas escuelas tampoco se aprecian grandes cambios en los resultados (Zabalza et. al., 1994).

Las acciones emprendidas, entre ellas el desarrollo de los PMEs, han contribuido a generar espacios de reflexión sobre las prácticas pedagógicas e intercambio de experiencias entre pares, siendo una actividad altamente valorada por los maestros.

No obstante, los profesores reconocen problemas para su realización, derivados principalmente de la distribución del tiempo al interior de las escuelas, lo cual dificulta la posibilidad de aunar los tiempos de los distintos maestros (Cerda et al., 1996).

Los profesores atribuyen el incremento de los resultados advertido en el SIMCE a factores como las nuevas estrategias empleadas, los materiales didácticos incorporados al proceso de enseñanza y aprendizaje y a la evaluación sistemática en el aula. Por otra parte, los bajos logros atribuidos a los alumnos (problemas de aprendizaje, falta de estudio, bajo desarrollo de las habilidades cognoscitivas), a las familias por su modelo cultural inadecuado y a la gestión escolar.

Impacto social del SIMCE sobre las familias

El impacto sobre las familias puede analizarse desde dos puntos de vista. Uno se refiere a la medida en que los resultados influyen en la elección del colegio para sus hijos y, el otro, se vincula con el incremento de la influencia y participación de los padres en la escuelas.

No fue posible encontrar información sistemática sobre el primer aspecto. No obstante, las observaciones informales permiten sostener que los padres han ido incorporando, gradualmente, la consulta de los resultados del SIMCE cuando se trata de seleccionar el colegio al que enviarán a sus hijos, sobre todo desde que dicha información ha recibido amplia cobertura en los medios de comunicación. Desde luego, este uso se produce entre los padres de los sectores sociales medios y altos, los que efectivamente pueden seleccionar entre los colegios públicos y privados gratuitos o pagados. En el caso de los padres de sectores pobres, dichas opciones no existen, porque no tienen más alternativa que enviar a sus hijos a escuelas gratuitas cercanas al lugar donde viven y que, en los sectores desfavorecidos, son muy pocas.

Para el análisis del segundo efecto, tampoco se ha encontrado evidencia de que el conocimiento de los resultados del SIMCE se haya traducido directamente en una mayor participación e influencia de los padres en el proceso educativo. Sin embargo, en una evaluación de los PMEs (Cerda et al., 1996) se encontró que, cuando en las escuelas involucraban a los padres en actividades pedagógicas específicas en el marco del proyecto y si a través de esta interacción pedagógica con sus hijos se daban cuenta de que estos últimos aprendían nuevas habilidades, entonces los padres comenzaban a participar más en la escuela.

A través de estos ejemplos, se ha procurado ilustrar el uso conceptual (análisis de resultados sin mayores consecuencias) e instrumental (análisis, seguido de acciones específicas) del SIMCE. Como síntesis se puede plantear que, pareciera que todavía las políticas ejecutadas a partir de los sistemas de medición desde el MINEDUC, presentan un impacto mucho mayor que las iniciativas emprendidas a nivel de las escuelas de forma autónoma, ya que éstas se ven aún bastante limitadas. Este hecho

puede ser un resabio del alto grado de centralización del sistema, que redunda en que las escuelas aún esperen las instrucciones desde el nivel central para emprender cualquier tipo de acciones.

Efecto del Sistema de Ingreso a la Educación Superior sobre las decisiones de los estudiantes

El impacto del Sistema de Ingreso a la Educación Superior sobre las decisiones de los estudiantes ha cambiado a lo largo del tiempo, debido en buena medida al cambio de estructura que ha experimentado este subsistema educativo.

Efectivamente, hasta 1981 el sistema de educación superior ofrecía a los egresados de la educación secundaria la oportunidad de ingresar a una de las universidades, unas pocas oportunidades de estudios no universitarios de muy bajo prestigio, o bien, ingresar al mercado laboral. En la práctica, la mayor parte de los egresados que no lograba de inmediato el ingreso a la universidad, optaba por repetir varias veces los exámenes de ingreso. Con frecuencia, después de varias tentativas[2], lograban ingresar a la universidad, pero no necesariamente al programa de su preferencia y, en general, a una carrera de menor prestigio social (Himmel y Maltes, 1980; Himmel y Maltes, 1987). Este proceso afectaba a las familias y al estudiante en términos de tiempo, recursos económicos y frustración.

Con la reestructuración de la educación superior, un número importante de estudiantes de los sectores socioeconómicos medio y alto han encontrado una solución en las universidades privadas, las cuales, en general, ofrecen las carreras de alto prestigio social y de buenas oportunidades laborales (Muga y Rojas, 1993).

Por otro lado, los egresados de la educación secundaria, provenientes del sector socioeconómico bajo, no pueden acceder a las universidades privadas y encuentran una solución muy parcial en las instituciones de educación superior no universitarias.

En suma, las pruebas de ingreso siguen teniendo un impacto sobre una decisión tan vital de los jóvenes como lo es la de su futuro profesional.

Efectos inesperados de los sistemas de evaluación del rendimiento escolar

A continuación se discutirán algunos ejemplos que introducen distorsiones en los propios sistemas, así como también efectos que repercuten sobre los estudiantes y sus familias.

Uso de los resultados del Sistema de Ingreso a la Educación Superior en las políticas de financiamiento

El Aporte Fiscal Indirecto (AFI) consiste en una política de financiamiento estatal

para las instituciones de educación superior, basada en los resultados de la PAA. Hasta los comienzos de la década de los años ochenta el sistema universitario chileno se encontraba financiado casi en su totalidad por el Estado, ya que los derechos de matrícula y aranceles eran prácticamente simbólicos y los ingresos provenientes de la venta de servicios y donaciones representaban una proporción ínfima dentro del financiamiento de las instituciones. Los aportes financieros del Estado se asignaban de acuerdo al criterio de "participación histórica" de las instituciones de educación superior y a la capacidad de cada entidad para negociar mayores aportes. Esta forma de asignar los recursos contribuía a un alto nivel de inestabilidad en las instituciones y a una planificación muy precaria de sus actividades [3].

No existía tampoco, en esa época, ningún mecanismo de control de gestión, directo o indirecto, que pudiera evaluar el rendimiento que generaban esos recursos, los cuales, a pesar de haber disminuido en cuanto a participación relativa en el presupuesto fiscal en educación, aún representaban un porcentaje alto del mismo (Lehmann, 1993).

A partir de la creciente necesidad de definir una política de financiamiento estable y como una forma de establecer ciertos mecanismos de control indirecto, se dicta en 1981 un Decreto con Fuerza de Ley que establece tres modalidades para el traspaso de los recursos:

- "Un aporte fiscal directo (AFD) destinado a financiar la investigación científica y tecnológica, así como la extensión universitaria.
- Un aporte fiscal indirecto (AFI) que se distribuiría entre las instituciones acreditadas, según el número de los 20.000 mejores postulantes que ingresaran en ellas.
- Crédito Fiscal Universitario (CFU) destinado a financiar a aquellos estudiantes que no disponen de recursos para pagar de inmediato el beneficio privado de la enseñanza superior, el cual sería reintegrado a largo plazo y con bajo interés" (Lehmann, 1993).

De estas tres modalidades interesa, para los efectos de este ejemplo, el AFI, pues la misma ley explicita que los veinte mil mejores alumnos son "aquéllos que han logrado un puntaje en la PAA que los ubique dentro de los veinte mil primeros lugares de la lista de alumnos que hayan rendido dicha prueba para el año respectivo y se matriculen en las universidades o entidades de educación superior..." (citado por Grez y otros, 1993: 1). Con esta definición se dio acceso a este fondo, a las universidades privadas sin AFI.

Hasta la promulgación de esta ley, los puntajes de la PAA constituían el único criterio común con una escala de interpretación comparable para identificar a los mejores alumnos, pues las calificaciones de la educación secundaria que era el otro criterio posible, presentan los conocidos problemas de comparabilidad entre las escuelas. Además, las calificaciones tienen una baja capacidad para diferenciar entre

los alumnos, pues presentan una distribución extremadamente asimétrica. Este último problema deriva del hecho de que los profesores, sabiendo que las calificaciones son un criterio en la selección universitaria, tratan de favorecer a los estudiantes con calificaciones altas[4], sobre todo en los últimos años de la educación secundaria y, por otro lado, reciben muchas presiones de los padres.

Esta política ha influido en el proceso de selección en lo que se refiere a la manipulación de los fondos que se otorgan a la PAA en el proceso de selección. En efecto, el coeficiente de ponderación de la PAA subió más allá de los límites académicamente recomendables, pasando de un 42,9 por ciento en 1979 en promedio a un 63,9 por ciento en 1982, valor que aproximadamente se mantiene hasta ahora (Grez et al., 1995).

El cambio en la política universitaria afecta a los alumnos de los sectores desfavorecidos en sus posibilidades de acceso a las universidades que disponen de crédito universitario. Estas instituciones reciben a los alumnos con los mejores puntajes en la PAA, pero dichos alumnos provienen mayoritariamente de sectores medios y altos (Himmel y Maltes, 1981).

Podría pensarse que se trata de un ejemplo de utilización de información evaluativa, que se inscribe en el uso simbólico o persuasivo. No obstante, de hecho también se puede considerar como instrumental. Si bien esta forma de empleo no estuvo prevista en los objetivos originales del sistema de ingreso, constituye un indicador de la efectividad de las instituciones de educación superior para reclutar a los mejores alumnos y, sin duda, tiene una relación con el prestigio de la formación que ofrecen estas entidades (Himmel y Maltes, 1980). Además, las instituciones privadas de educación superior, publican el monto de los aportes que reciben a través de este canal.

Instituciones que preparan para las pruebas de ingreso a la educación superior

Con la puesta en marcha del sistema de ingreso a la educación superior, comenzaron a surgir instituciones privadas llamadas "preuniversitarios" que preparan a los estudiantes para las pruebas de ingreso. El número de estas instituciones ha crecido enormemente y, en la actualidad, más del 50 por ciento de los candidatos se prepara en los preuniversitarios. Los colegios también han puesto en marcha programas paralelos de preparación para los exámenes de selección. Estas iniciativas implican un gasto adicional para los padres y, por lo tanto, sólo acceden a ellas los alumnos que disponen de medios para hacerlo. No obstante, no está en absoluto probado que la asistencia a un preuniversitario contribuya a incrementar significativamente los puntajes (Rojas, 1985).

Aún más, muchas de estas instituciones realizan una amplia difusión de los altos puntajes que alcanzan los estudiantes inscritos en sus cursos, pero se ha detectado al mismo tiempo que sólo aceptan candidatos que tienen muy buenas calificaciones en la escuela secundaria. De este modo, los preuniversitarios hacen un uso

simbólico de los resultados del sistema.

Uso de los resultados de la PAA como requisito para postular a un trabajo

Desde hace varios años es frecuente que en los periódicos se publiquen anuncios con ofertas de trabajo, tales como estafeta, ascensorista, chofer, mensajero y otros trabajos menores, que formulan como requisito para postular a ellos, el tener la PAA rendida.

Distorsiones introducidas por los directores de escuela en el SIMCE

En los últimos años se detectó que algunos directores de escuelas inducían a los alumnos de bajas calificaciones a no asistir durante los días de aplicación de las pruebas del SIMCE, con el fin de obtener rendimientos más altos en la escuela. Frente a esta constatación, se decidió no proporcionar los informes de resultados a estas escuelas (Prado, 1995)[5]. Algunos datos informales revelan que los colegios que habían recurrido a la estrategia mencionada, al año siguiente presentaban una asistencia normal. Sin embargo, el problema no ha desaparecido a nivel de la población de escuelas.

En este caso, también se puede señalar que la utilización que se ha hecho de los resultados, así como el uso futuro que se piensa hacer de ellos, se ubican en la categoría de uso simbólico.

Todas los ejemplos descritos muestran el uso simbólico e instrumental de los resultados que, si bien se encuentra seguido de acciones, no responde a los propósitos originales del sistema de evaluación respectivo.

En relación a efectos inesperados de los sistemas de evaluación, sólo cabe agregar que son difíciles de predecir, pues en cada contexto pueden surgir efectos diferentes. No obstante, es factible que una buena parte de los efectos inesperados surja de los sectores afectados por el sistema, ya sea positiva o negativamente.

Factores que han favorecido o inhibido el impacto social de los sistemas de evaluación del rendimiento escolar

Al comienzo de esta exposición, se señaló que un sistema de evaluación podía tener impacto social cuando sus resultados se utilizasen en la toma de decisiones conducentes a acciones que afecten a diversos grupos sociales. Consecuentemente, los resultados de las evaluaciones sólo pueden tener efectos si son usados y, en mayor medida, si el uso es conceptual. Por este motivo, se analizarán los factores que favorecen o dificultan el uso de las evaluaciones y se procurará ejemplificarlos con aspectos relevantes al caso chileno.

Los factores suelen clasificarse en tres grupos: humanos, contextuales y evaluativos

(Alkin, 1985). Los factores humanos constituyen características de los evaluadores y usuarios de la evaluación. Los condicionantes contextuales se refieren a los obstáculos preexistentes a la evaluación y los factores evaluativos relacionados con las características propias del sistema de evaluación. Todos estos factores se relacionan entre sí y para cada grupo es posible reconocer dimensiones más específicas (Figura 3).

Figura 3. Factores condicionantes del uso evaluativo

Factores humanos

Entre las características de los evaluadores, que promueven el uso de la información evaluativa, se encuentra la credibilidad, esto es la confianza que inspiran los evaluadores a los usuarios. Esta característica tiene que ver con la experiencia, la capacidad técnica y el reconocimiento que tienen entre los usuarios. El equipo profesional que puso en marcha el PER y el SIMCE tenía a su favor esta característica, por cuanto sus miembros eran externos al MINEDUC y se desempeñaban en una universidad de prestigio. Asimismo, el equipo que desarrolló el sistema de selección universitaria se había acreditado por su trabajo de varios años previos a la implantación del sistema.

Una segunda característica importante de los evaluadores es la sensibilidad política. El equipo del PER tuvo algunas limitaciones respecto a esta característica. De hecho, no fue capaz de comprender oportunamente el conjunto de los procesos políticos que rodeaban el PER, ignorando, sobre todo, la falta de consenso político,

pues confiaba en demasía en sus capacidades técnicas.

En relación a las características de los usuarios se encuentra el compromiso que éstos tengan con el uso de la información, que se encuentra determinado, entre otras cosas, por lo necesaria que consideren la realización de la evaluación. En el caso del PER, los sucesivos equipos ministeriales, que no habían estado en la iniciación del programa, no sentían que el PER era necesario y, por tanto, algunos llegaron a aceptarlo, pero de hecho no tenían mayor compromiso con su uso.

Otra característica de los usuarios que compromete su uso es el estilo profesional, que está relacionada con la existencia de una "cultura evaluativa" en el medio educativo. La "cultura evaluativa" implica la posesión de competencias profesionales en evaluación educativa, el liderazgo organizativo y el uso de la información producto de la evaluación en la labor cotidiana. En el caso del SIMCE y del sistema de ingreso a las universidades esta "cultura evaluativa" se ha ido desarrollando gradualmente en el tiempo, lo que ha contribuido a consolidar ambos sistemas.

Factores contextuales

Uno de los obstáculos que puede atentar contra el uso de la información evaluativa es la estructura organizativa en el que se inserte el sistema de evaluación. En este sentido, cuando la entidad ejecutora de la evaluación es externa y no identifica claramente cúales son las limitaciones de la estructura organizativa, puede encontrarse con serias resistencias al uso de la información. Así sucedió, por ejemplo, durante la puesta en marcha del PER, pues no se comprendió que dentro del MINEDUC había un sector influyente que se consideraba perfectamente capaz de llevar a cabo el proyecto y que opuso una resistencia tenaz a la ejecución del mismo. En cambio, durante la primera etapa del SIMCE, se incorporó a una parte significativa de los miembros de dicho sector, los cuales se harían cargo del sistema al terminar dicha etapa; fue de este modo que se superó gradualmente esta dificultad.

El factor precedente se puede calificar como intraorganizacional, mas también se pueden considerar dimensiones extraorganizacionales, que a su vez pueden favorecer o inhibir el uso de la información evaluativa. Entre estos factores se encuentra el consenso político en torno de la necesidad de realizar la evaluación. La falta de consenso político puede poner en riesgo, entre otras cosas, el financiamiento estable del sistema (Himmel, 1996). Este consenso no tuvo lugar respecto al PER. En cambio existe desde hace muchos años en relación al sistema de ingreso a las universidades y al SIMCE.

Factores evaluativos

En esta categoría se inscriben factores tales como los procedimientos utilizados para la evaluación, su rigor metodológico, las características de los informes evaluativos y la oportunidad con que son suministrados.

Tanto el PER, como el SIMCE y el sistema de ingreso a las universidades, han empleado procedimientos metodológicamente adecuados, aunque ninguno de ellos ha incluido aún un sistema de equivalencias entre sucesivas mediciones. Los informes en algunos casos no han sido lo suficientemente claros para el público a quien iban dirigidos, porque adolecían de un estilo demasiado técnico. En general, la información ha sido oportuna en los sistemas chilenos y, si por alguna causa se ha suministrado con retraso, entonces los usuarios han puesto de manifiesto esta circunstancia, señalando que no ha podido ser utilizada por dicho retraso (Zabalza, 1994).

Recomendaciones para políticas sobre sistemas nacionales de evaluación del rendimiento escolar

En primer lugar, es necesario decidir si el sistema de evaluación debe incorporarse al educativo. En este sentido, cabe preguntarse si las condiciones del país lo permiten.

Entre las condiciones necesarias para la implantación de los sistemas nacionales de evaluación, se encuentra la disponibilidad de recursos humanos y materiales que se puedan destinar a esta tarea. Éstos requieren de una cantidad significativa de recursos financieros y, por ello, es necesario evaluar si dichos recursos pueden ser utilizados para tal propósito sin perjudicar otras inversiones del sistema educativo que pueden ser prioritarias. Por otro lado, la disponibilidad de recursos humanos para llevar a cabo un sistema nacional de evaluación tiene que estar asegurada, pues se trata de procesos altamente técnicos. Si no se dispone inmediatamente de los especialistas es necesario tener en cuenta que la formación de un equipo profesional competente demora varios años y que esta misma circunstancia puede encarecer la puesta en marcha del sistema.

Otra condición que merece ser analizada es el consenso político en torno a la necesidad de ejecutar un sistema de evaluación y, si éste existiera, con qué objetivos se pondría en marcha. Como ya se señaló, un sistema sólo puede responder a un número relativamente pequeño de propósitos.

Definidos los objetivos, es indispensable analizar para qué sectores del sistema educativo tendrá un impacto social. Por ejemplo, supóngase que el sistema tiene la intención de distribuir a los alumnos que pasan de la educación primaria a la secundaria entre una modalidad orientada a proseguir estudios superiores y otra, a la inserción laboral. En este caso, el sistema tiene consecuencias cruciales para los estudiantes y sus padres y bastante más indirectas para los maestros y directores de escuela. Por otra parte, si el sistema tiene como finalidad la de evaluar políticas educativas, entonces el impacto sobre los alumnos y sus familias es relativamente escaso, mientras que conlleva serias consecuencias para las autoridades educativas, los profesores y los directores de escuela. Como puede advertirse, el cambio en el objetivo modifica substancialmente el grupo social que es afectado por el sistema de evaluación. Con estos ejemplos se quiere destacar la importancia de identificar a los

grupos más afectados, en los que es plausible que se generen las mayores resistencias y los efectos inesperados.

Otro asunto que debe resolverse es el referente a qué organismo operará el sistema. Se pueden plantear un conjunto de opciones que van desde la ejecución llevada a cabo por un organismo público, eventualmente inserto en el Ministerio de Educación, hasta realizada por un organismo privado. Si la tarea es desarrollada por un organismo del Ministerio de Educación, tiene la desventaja de hacer aparecer a esta secretaría de Estado como "juez y parte en la evaluación". Además, resulta particularmente difícil en América Latina mantener a un equipo profesional altamente calificado dentro de la administración pública—por cuanto las remuneraciones del sector público son extremadamente bajas. Esto produce una elevada tasa de rotación de personal calificado, que resulta perjudicial en un proyecto de esta naturaleza. Este esquema probablemente presente ventajas en cuanto al costo y a las posibles resistencias en el sistema educativo.

Por otra parte, si un organismo externo es el ejecutor, el sistema puede verse muy afectado por las resistencias intraorganizacionales del Ministerio y, además, el sistema de evaluación resulta, seguramente, más caro. No hay que olvidar, sin embargo, que tiene la ventaja de ser independiente, con lo cual aumenta su credibilidad.

Otra opción estaría en la posibilidad de crear un organismo técnico, no gubernamental y externo que, bajo la supervisión del Ministerio de Educación, se haga cargo del sistema. A ello se puede agregar un Comité Nacional, con una amplia base de representación de diferentes organizaciones sociales, que defina las políticas del sistema de evaluación para cada período.

Una vez que exista la convicción de poner en marcha el sistema, se hayan acordado sus intenciones y definido quién estará a su cargo, es indispensable abordar los rasgos específicos que tendrá el sistema. Entre estos aspectos se pueden mencionar los siguientes:

- Cobertura del sistema: censal o muestral. Se trata de optar entre aplicar el sistema a toda la población estudiantil definida, o bien, a una muestra. Aunque en esta decisión juega un papel preponderante el objetivo del sistema, en ocasiones es necesario considerar otros aspectos adicionales. Ambas posibilidades tienen sus ventajas y limitaciones. Si la población escolar es muy grande, es prácticamente imposible que un país invierta los recursos necesarios para financiar el sistema. Empero, la cobertura censal alberga el potencial de facilitar a cada interesado en la educación con la información más pertinente, lo que incrementa la posibilidad de su uso. Por su parte, la cobertura muestral presenta la ventaja de tener menor costo, pero la información proporcionada puede ser juzgada como menos congruente con las expectativas de los usuarios, lo que a su vez reduce la probabilidad de su utilización.
- Alcance curricular. Como se trata de evaluar el rendimiento escolar , es inelud-

ible abordar , al menos, tres áreas. La primera, corresponde a los grados en los que se practicará la evaluación, ya que es muy poco probable que se aborden todos y por lo tanto este tema exige la definición de cortes en los planes de estudios que sean significativos para el proceso de escolarización. Respecto a la segunda área, concomitante a la anterior, exige especificar si las mediciones abarcarán sólo las competencias, el contenido o los objetivos de los grados especificados, o si incluirá los logros de la etapa que los precede. Por último, es fundamental determinar las áreas del currículo que se incluirán, pues no se puede abarcarlas todas, bien sea por el tiempo que requieren, la viabilidad técnica y/o el costo.

- Periodicidad con que se aplicará el sistema de evaluación. El esfuerzo que requiere poner en marcha un sistema de evaluación exige que tenga continuidad. Por tal razón, es esencial que se asegure su periodicidad de modo que proporcione regularmente las informaciones requeridas. Si el sistema está destinado a certificar los logros de aprendizaje de los alumnos, a seleccionarlos o distribuirlos, esto lleva consigo decisiones anuales acerca de ellos, forzosamente deberá ser anual. En cambio, si tiene por objetivo evaluar políticas educativas, realizar un seguimiento o responsabilizar a los educadores por los aprendizajes logrados por los alumnos, el período que transcurre entre las sucesivas aplicaciones, deberá ser necesariamente mayor, pues los cambios demoran al implicar la movilización de un conjunto de agentes muy numeroso.
  Además, para determinar la periodicidad, conviene considerar los grados que abarcará el sistema de evaluación, especialmente en el caso de que comprenda un seguimiento, a fin de afianzar la posibilidad de efectuarlo. Por ejemplo, si se evalúa al finalizar el tercero y sexto cursos, es aconsejable que la periodicidad permita evaluar la misma cohorte que fue evaluada en tercero, y sexto cursos.
- Vinculación con otros sistemas de información. Los sistemas nacionales de evaluación son a la vez sistemas de información y no son los únicos que existen en el medio educativo. En este sentido, es conveniente articularlos desde el comienzo con los otros sistemas de información educativos y sociales para evitar las redundancias en la recolección de los datos y permitir el intercambio de información con los demás sistemas.

En este estudio se han analizado algunas de las áreas problemáticas concernientes al impacto social de los sistemas nacionales de evaluación del rendimiento escolar, desde la experiencia de los sistemas chilenos. Este análisis es, ciertamente, parcial y sesgado. Sin embargo, se espera que pueda aportar algunos antecedentes útiles, teniendo siempre presente que los impactos sociales, más específicos en cada país, ya sean esperados o inesperados, sólo pueden desprenderse del contexto particular de cada caso.

NOTAS

[1] Las escuelas privadas gratuitas reciben una subvención por alumno del Estado, la que cubre

su financiamiento total o parcial.

[2] Para tener derecho de postular a las universidades financiadas por el Estado, los candidatos deben lograr por lo menos 450 puntos en la PAA. Un 38 por ciento de los estudiantes no logran este puntaje. De este modo, se informa que algunos alumnos repiten las pruebas hasta siete veces. Al grupo anterior se agregan todos aquéllos que lograron ingresar a un programa que no era de su preferencia y quieren cambiar de programa (Rojas, 1984). En Chile, hasta hace unos pocos años, los estudios universitarios tenían muy poca flexibilidad para transferencias entre programas de estudios. Esto implicaba que un estudiante tenía que volver a rendir los exámenes de ingreso si quería cambiar de programa. Más difícil aún era el traslado de una universidad a otra.

[3] Hasta 1980 existían en Chile ocho universidades nacionales, esto es, con su sede central en una ciudad y, subsedes en otras. De éstas, dos eran estatales y seis, privadas, financiadas por el Estado. La reforma de 1980 reestructuró este sistema, de modo que ninguna de las universidades originales pudo conservar sus subsedes. Es así como las ocho instituciones originales se transformaron en 23 universidades, de las cuales 14 son estatales y nueve, privadas con subsidio estatal, acceso al AFI y al crédito universitario. Además, durante el período 1981-1995, se fundaron 44 universidades privadas, que sólo tienen acceso al AFI (Muga y Rojas, 1993). A las universidades, se suma un alto número de Institutos Profesionales y Centros de Formación Técnica, tanto públicos como privados, creados desde 1981 y que han ido ganando en prestigio entre los padres y alumnos.

[4] Este fenómeno ha sido comprobado porque entre 1967 y 1996 el promedio nacional de las calificaciones brutas ha subido de 4,7 a 5,6 en una escala de 1 a 7, mientras tanto, los desempeños en las pruebas de conocimientos se han mantenido prácticamente invariables en todos estos años.

[5] Como la matrícula del curso se incorpora tres meses antes de la aplicación de las pruebas a las bases de datos del SIMCE, la comparación entre esta información y la asistencia el día de las pruebas, reveló diferencias mayores que el promedio, en un conjunto de colegios. A raíz de datos, se optó por no proporcionar los informes de resultados a dichas escuelas. Esto implica también que en las publicaciones del periódico figuren con la información en blanco.

REFERENCIAS

Alkin, M. (1981). "The Feasibility of Measuring Educational Attainment in Chilean Schools." Consultant Report. Santiago de Chile: PER, Informe Nº8.

Alkin, M. C. (1985). *A Guide for Evaluation Decision Makers*. Beverly Hills, California: Sage Publications.

Alkin, M. C.; Daillak, R. Y White, P. (1979*). Using Evaluations: Does Evaluation make a difference?* Beverly Hills, California: Sage Publications.

Braskamp, A. L. (1982). "A definition of use." *Studies in Educational Evaluation*. 8, pp. 169-174.

Brown, R. D.; Newman, D. y Rivers L. (1985). "An exploratory study of contextual factors as influences on schoolboard evaluation information needs for decision

making." *Educational Evaluation and Policy Analysis*. 7: 437-445.

Cabezón, E., Condemarín, M y Vaccaro, L. (1996). "Jóvenes monitores: una esperanza de recuperación del ethos normalista." Santiago, Chile: *Revista de Educación*, 240: 50-55.

Cerda, A. M.; Assaél, J. y Rodas, M.T. (1996). "Una mirada evaluativa a los Proyectos de Mejoramiento Educativo." Santiago, Chile: *Revista de Educación*, Nº 236: 51-55.

Contreras, J. (1988). "Impacto del Programa de Evaluación del Rendimiento Escolar (PER) a nivel institucional: Establecimientos Educacionales." Tesis para optar al grado de Magister en Ciencias de la Educación. Santiago, Chile: Pontificia Universidad Católica, Facultad de Educación.

Davis, G. A. y Thomas, M. A. (1989). *Escuelas eficaces y profesores efecientes.* Madrid: La Muralla.

Enrione, A. E. (1991). "Liderazgo del Director." Tesis para optar al grado de Magister en Ciencias de la Ingeniería, Pontificia Universidad Católica de Chile, Escuela de Ingeniería.

Greany, V. y Kellaghan, Th. (1996). *Monitoring Learning Outcomes of Education Systems.* Washington, DC: Directions in Development, The World Bank.

Grez, N., Cazenave, J., González, M. y Gil, F.J. (1993). «Una propuesta al Proceso de Selección de las Universidades Chilenas. Documento de Trabajo Nº7/93.» Santiago de Chile: Corporación de Promoción Universitaria.

Himmel, E. (1996). "National Assessment in Chile." En Murphy, P; Greany, V.; Lockheed, M y Rojas, C. Eds. *National Assessments. Testing the System.* Washington, DC: EDI Learning Resources Series, The World Bank: 111-128.

Himmel, E., Majluf, N., Maltes, S. y Gutiérrez, J. (1982). "Análisis de los efectos de Variables Socioecomómicas y del Colegio sobre el Rendimiento Escolar." Santiago de Chile: Pontificia Universidad Católica de Chile, Proyecto DIUC Nº42/81, Informe Final.

Himmel, E., Majluf, N., Maltes, S., Gazmuri, P, Contreras y Escobar C. (1988). "Análisis del Impacto sobre el Sistema Educacional de 3 años de aplicación del Programa de Evaluación del Rendimiento Escolar.» Proyecto FONDECYT. Informe Final.

Himmel, E. y Maltes, S. (1980). "El prestigio de las carreras universitarias en Chile" Santiago, Chile. *Revista de Educación*: 33-37.

Himmel, E. y Maltes, S. (1987). "Factores que inciden en la decisión profesional y el rendimiento académico." En *Orientación: Nuevas tendencias y perspectivas*. Santiago de Chile: Ediciones Universidad Católica de Chile: 125-151.

Himmel, E., Majluf, N. y Maltes, S. (1980). "Efecto de Variables Macroestructurales y del Colegio sobre el Rendimiento en las Pruebas de Selección Universitaria." Santiago, Chile: Vicerrectoría Pontificia Universidad Católica de Chile, Vicerrectoría Académica, Informe Nº10.

Himmel, E. y Maltes, S. (1981). "¿Quiénes son los 20.000 primeros puntajes de la Prueba de Aptitud Académica?" Santiago, Chile: Pontificia Universidad Católica de Chile, Vicerrectoría Académica, Informe Nº2.

King, J. A. y Peachman, E. M. (1984). Pinning a wave to the shore: conceptualizing evaluation use in school system. *Educational Evaluation and Policy Analysis*, 6:

241-251.

Lara, F. (1985). "Ingreso a la Educación Superior: El Sistema Nacional." Santiago de Chile: *Cuadernos del Consejo de Rectores de las Universidades Chilenas*, 24: 1-23.

Lehmann, C. (1993). "Financiamiento de la Educación Superior en Chile. Resultados del período 1982-1992." En *Informe de la Educación Superior 1993*. Santiago de Chile: Colección Foro de la Educación Superior: 127-176.

Lockheed, M. (1996). "International Context for Assessments." En Murphy, P; Greany, V.; Lockheed, M. y Rojas, C. Eds. *National Assessments. Testing the System.* Washington, DC: EDI Learning Resources Series, The World Bank: 9-19.

Lockheed, M.E. (1991). "World Bank support for capacity building: The challenge of educational assessment." Trabajo presentado al International Consultive Forum, París.

Ministerio de Educación. (1997). "Resultados SIMCE 1996: ¿Qué pasó con los Aprendizajes?" Santiago de Chile: *Revista de Educación*, 244: 20-23.

Muga, A. y Rojas, F. (1993). "Estadísticas de la Evolución del Sistema de Educación Superior en Chile: 1980 - 1992". En *Informe de la Educación Superior 1993*. Santiago de Chile: Colección Foro de la Educación Superior: 39-126.

Parraguez, W. (1992). "Profesores Rurales Rompen el aislamiento." Santiago, Chile: *Revista de Educación*, 198: 16-20.

Prado, M. N. (1995). "Un SIMCE no se puede improvisar". Santiago, Chile: *Revista de Educación*, 198: 9-11.

Rojas, R. (1983). "Antecedentes de los rezagados y actividades realizadas por ellos entre el egreso de la enseñanza media y 1983". Santiago, Chile: Universidad de Chile, Dirección General Académica y Estudiantil, Monografía 19.

Rutherford, W L. (1985). "School principals as effective leaders." *Phi Delta Kappa*, 67: 31-34.

Schiefelbein, E. (1992). "Análisis del SIMCE y sugerencias para mejorar su impacto en la calidad." En Gómez, S. Ed., *La Realidad en Cifras. Estadísticas Sociales.* Santiago de Chile: FLACSO/INE/UNRISD: 241-280.

Zabalza, J.; Rojas, V. y Soto, J. (1994). "Diferencias en el Rendimiento Académico en el SIMCE, en Sectores Socioeconómicos Medio-Bajos y los Principales Factores Explicativos Atribuibles al Establecimiento Escolar." Proyecto FONDECYT 920742. Informe Final.

CAPÍTULO 6

SISTEMA DE EVALUACIÓN DE LA CALIDAD DE LA EDUCACIÓN EN COLOMBIA

Gabriel Restrepo[1]

Este trabajo, al igual que el anterior, describe una experiencia nacional de tres décadas en la aplicación y el análisis de pruebas de aprendizaje escolar. El sistema nacional de pruebas de Colombia ha demostrado una gran estabilidad y capacidad gerencial y goza de la confianza de las instituciones y los usuarios. Las pruebas se utilizaron inicialmente con propósitos selectivos y posteriormente para mejorar la calidad de la educación. El caso analiza el contexto global y nacional de la transición y los mecanismos que se establecieron para responder a nuevas necesidades impuestas por las recientes reformas de la educación.

INTRODUCCIÓN

Problemas y dilemas

Los sistemas educativos de diferentes países afrontan hoy una especie de paradoja, derivada de dos tendencias al parecer contradictorias. La primera, el advenimiento de la aldea global (McLuhan, 1985: 45). La segunda, la nueva valoración de la vida local. La globalización de los mercados y la "convergencia tecnológica" (Nelson y Wright, 1992) fuerzan a los individuos a desarrollar competencias universales, las cuales son hasta cierto punto indiferentes a espacios o a tiempos particulares. Provienen las mismas de los requerimientos científico-técnicos contemporáneos que hoy anulan las antiguas ventajas comparativas. Configuran una especie de ciudadanía universal, ya no inducida por la religión o por la ideología, sino por la racionalidad instrumental. Racionalidad que, sin embargo, es el mito por excelencia en el mundo contemporáneo (Alexander, 1991: 283-309). A tal educación abierta apuntan la extensión y velocidad de los medios de comunicación y las redes internacionales de información.

Por otra parte, la crisis de legitimidad de los Estados y la deconstrucción de los

discursos del poder (Weiler, 1992), ponen en tela de juicio una educación basada en racionalidades impuestas desde afuera. A partir de una valoración del mundo de la vida cotidiana (Habermas, 1987), muchos reclaman una educación fundada en estilos de conocimiento locales y privilegiando la autoevaluación sobre la evaluación externa, los estudios "cualitativos" sobre los "cuantitativos", el afecto sobre el entendimiento y el valor absoluto del individuo sobre las agregaciones.

Tensionados entre ideales universalistas y locales, los sistemas de educación experimentan una crisis que unas veces se resuelve en parálisis, otras en incertidumbre y, en casos afortunados, en creatividad. La escuela, el colegio o la universidad han dejado de ser instituciones autosuficientes, como eran antes de la revolución científico-técnica, para convertirse en piezas de un engranaje más complejo de reproducción de la misma sociedad (Coombs, 1971). Así, las políticas de los Estados han de atender al mismo tiempo dos impulsos que parecen contrapuestos: integrar a los ciudadanos en redes cada vez más internacionales y conceder mayor autonomía a los individuos y a las comunidades locales.

La tensión se extiende a los sistemas nacionales de evaluación. ¿Deben éstos responder a exigencias universales o a necesidades locales?; ¿deben limitarse a registrar saberes y competencias producidas en contextos diferentes, bajo el supuesto de que ellas tienen valor inconmensurable? Dado que la evaluación es "el poder de determinar el valor de algo" (Kvale,1992: 119), ésta oscila entre el universalismo de la racionalidad instrumental y los particularismos de las preferencias locales, con sus sesgos subjetivos. La descentralización y evaluación son una aspiración deseable, pero podrían entrar en contradicción si no se hace un esfuerzo por conciliarlas (Kogan, 1992).

Las tensiones se viven en forma más dramática en sociedades que no han accedido a una economía, a un estado o a una cultura en el sentido moderno, sociedades en las cuales—dada la interrogación abierta por el posmodernismo—la crisis mencionada puede resolverse de modo fácil en la parálisis, cuando no en la mera perplejidad.

Propósito del ensayo: el caso de Colombia

En este ensayo se expone el esfuerzo de Colombia por construir un sistema nacional de evaluación de la calidad de la educación que, sin escapar a las tensiones mencionadas, muestra una continuidad en treinta años y mucha creatividad para superarlas.

En 1968, Colombia creó un sistema nacional de selección para el acceso a la educación superior, a partir del cual se inició en 1990 un sistema de evaluación nacional de la calidad de la educación. No es fácil el paso de un sistema de selección a otro de evaluación, porque son modelos distintos (Greagney y Kellaghan, 1996: 31-39; Rodríguez, 1982: 35-38). No obstante, lejos de ser antagónicos, pueden ser complementarios. La experiencia de Colombia demuestra que es posible extraer las mejores ventajas de uno y otro modelo.

En una primera parte de este ensayo se ofrece una síntesis de la evolución de la educación (1950-1990), en la cual se subrayan las limitaciones y a la vez las oportunidades para el desarrollo de un sistema nacional de selección para el acceso a la educación superior. En la segunda parte se examina el paso de un sistema de selección a uno más amplio de evaluación (1990- 1997), en el marco de la evolución más reciente de la educación y la sociedad. Se muestra entonces que dichos sistemas pueden ser complementarios. En la tercera parte se resumen las principales conclusiones y se exponen las opciones de política que puedan servir para otros países, según sus características específicas.

Para la redacción de este estudio se entrevistó a funcionarios públicos, directivos de universidades, colegios y escuelas, profesores, investigadores, estudiantes y familias. Se ha revisado un archivo de prensa con no menos de 800 entradas en los últimos 10 años. También se ha examinado la literatura existente sobre el tema. El Servicio Nacional de Pruebas ha proporcionado todo el apoyo en materia de información.

SISTEMA DE SELECCIÓN: LOS EXÁMENES DE ESTADO PARA EL INGRESO A LA EDUCACIÓN SUPERIOR (1964-1990)

La educación en Colombia hasta 1990

En Colombia, "la esperanza de educación se mantuvo sin crecimiento alguno en 1,4 años por persona desde comienzos del siglo hasta los años cincuenta" (Departamento Nacional de Planeación, 1997: 88). Hasta entonces, "Colombia era el país que registraba el mayor atraso relativo en educación en América Latina" (DNP, 1991: 88).

Sin embargo, a partir de 1950 y hasta 1988, "Colombia tuvo un crecimiento de su cobertura educativa mayor que cualquier país latinoamericano en el período— incluyendo Nicaragua— y tan sólo fue superada a nivel internacional por el Congo, Nepal y Togo" (DNP., 1991: 89-90).

Entre 1950 y 1990 las coberturas de educación primaria ascendieron del 45 por ciento a más del 90 por ciento; las de secundaria del 5 por ciento al 48 por ciento; la de educación superior del 1 por ciento al 12 por ciento (Duarte, 1997:1). Una voluntad política por superar la violencia entre los dos partidos tradicionales incidió en un aumento considerable del gasto público y privado en educación, que creció del 1 por ciento en 1950 al 3,5 por ciento en 1984 (Duarte, 1997: 1)

En 1953 se inició el planeamiento integral de la educación y se fundó un instituto de crédito educativo para estudios en el exterior, que sirvió para formar una masa crítica científica y técnica. Dicho instituto adoptó el mérito como criterio de selección. En 1960 el gobierno central asumió los costos de la educación primaria y en 1975 la financiación de la educación secundaria pública. En ambos casos introdujo controles, no siempre efectivos, sobre los gobiernos departamentales y

municipales (Duarte, 1995).

Pero en 1976 la expansión se detuvo. "Desde mediados de los años setenta el ritmo de crecimiento de la educación se ha reducido de una manera notable. En el nivel de primaria, las tasas de crecimiento anuales de la matrícula han sido inferiores al 1 por ciento. En la secundaria, en donde existen las mayores carencias, fue apenas del 5 por ciento entre 1975 y 1984, y del 2 por ciento entre 1985 y 1990. (...) La esperanza de educación dejó prácticamente de crecer en los últimos 15 años (...) hasta el punto de haber alcanzado en la década de los años ochenta una inequidad educativa mayor que la de todos los países del mundo, con excepción de la India" (DNP, 1991: 93).

El aumento en la participación de la educación en el producto interno bruto descendió desde 1984 y sólo una década más tarde recuperó la tendencia al alza, pese a que Colombia mantuvo un crecimiento positivo en la llamada década perdida (Figura 1).

Como resultado, las coberturas de la educación son aún hoy muy precarias: 20 de cada 100 niños entre 6 y 11 años están fuera de la escuela y la mitad de los niños y jóvenes entre 12 y 17 años no asisten a la secundaria. De quienes ingresan, muchos tienen poca probabilidad de mantenerse, debido a la ineficiencia de las instituciones educativas. Las diferencias son más agudas cuando se toman en cuenta las oportunidades individuales—menores para las regiones y los estratos pobres—y las ventajas institucionales—menores para el campo, ciudades pequeñas y establecimientos nocturnos; mayores, por lo general, para colegios privados o públicos en ciudades grandes (Duarte, 1997: 2).

El descenso del crecimiento de la educación observado desde 1976 se ha debido a imperfecciones del sistema educativo, que reflejan debilidades del sistema democrático. La expansión no obedeció siempre a un control racional pues, aunque el gobierno central asumió la mayor parte de los costos de la educación, no creó instrumentos efectivos de control. Por ejemplo, "sólo hasta 1991 tuvo datos más o menos confiables de cuántos docentes oficiales había y a qué nivel los pagaba" (Duarte, 1997: 8).

Los poderes locales manejaron la educación con criterio clientelista y con frecuentes prácticas corruptas (Gómez y Losada, 1984; Duarte, 1995). En los municipios, las transacciones individualistas predominaron sobre las negociaciones colectivas y normas universales, mientras que un gobierno central débil se oponía a un sindicato fuerte (Duarte, 1997: 8-10) y justificado a veces por las condiciones precarias del maestro.

La inestabilidad de los Ministros de Educación no favoreció el control racional: el promedio de duración de los Ministros de Educación ha sido de un año, a diferencia de los dos y medio de los de Hacienda. La inestabilidad ha sido aún mayor en los funcionarios departamentales y municipales, que rotan en función de intereses

políticos locales o regionales (Hansson, 1996).

Además, sólo hasta hace una década los economistas se han interesado de modo sostenido por la reflexión sobre los problemas de la equidad y distribución de los servicios sociales.

Antecedentes del sistema nacional de selección

Como se ha indicado, la cobertura de la educación superior pasó del 1 por ciento, en 1950, a cerca del 12 por ciento, en 1990, la cual equivale a medio millón de estudiantes más (Duarte, 1995: 60). En 1991, Colombia figuraba entre los países de desarrollo medio en América Latina en cobertura, inferior al 16,9 por ciento del promedio regional (Álvarez, 1995). Pero como el punto de partida fue tan bajo, las tasas de crecimiento anuales fueron elevadas. En 1989 había 236 instituciones de educación superior, 30 por ciento de las cuales eran oficiales y 70 por ciento privadas (DNP, 1991:103). En 1950, cuando no había más de 40 universidades, la proporción era la contraria.

Para regular la expansión universitaria, el gobierno creó en 1954 el Fondo Universitario Nacional. Por su parte, las universidades constituyeron en 1958 la Asociación Colombiana de Universidades. Dichas instituciones llevaron a cabo estudios sobre el bachiller colombiano, crearon servicios de orientación profesional y organizaron cuatro seminarios sobre admisión universitaria entre 1960 y 1966 (Acero, 1990: 34). En 1966 surgió el Servicio Nacional de Pruebas, adscrito a ambas entidades. El Servicio Nacional de Pruebas trazó una estrategia ambiciosa, que incluía las siguientes actividades (Acero: 33-41):

1. Pruebas de aptitud y vocacionales para el cuarto curso de bachillerato.
2. Pruebas preparatorias y de aptitud para el mejoramiento de la calidad de la enseñanza en colegios de educación secundaria.
3. Exámenes de conocimiento y aptitud para bachilleres egresados.
4. Creación de oficinas de admisión en cada universidad.
5. Inicio de un año universitario básico común por grandes áreas.
6. Pruebas selectivas al cabo de dicho año.
7. Pruebas para graduados universitarios.

Así, los fundadores pensaron en un sistema de selección que fuera parte de uno más amplio de evaluación de la calidad (Greagney y Kellaghan, 1996 : 31 y 39). El inicio del sistema de selección se debió ante todo al desarrollo de la psicología en Colombia, que ha sido excepcional (Ardila, 1973; 1993). Los fundadores del sistema se inspiraron en el Servicio de Pruebas Educativas (Educational Testing Service), donde en 1962 se entrenaron en la construcción del Examen de Aptitud Escolar (Scholastic Aptitude Test, SAT) usado por la Junta de Exámenes de Acceso a la Universidad (College Entrance Examination Board) (Caro, 1990: 3; Rodríguez, 1982: 7). Una versión del SAT se tradujo en Puerto Rico por la misma época, con participación de psicólogos colombianos. Luego, las pruebas se adaptaron a Colom-

bia con un sentido libre y creativo. Entre 1964 y 1967 el Servicio realizó pruebas graduales en algunas universidades públicas, con las que fue ganando confianza. Se acumuló experiencia por ensayo y error, en un proceso en el cual existía poca división de trabajo, mucha actividad manual, pero no poco entusiasmo.

Hecho decisivo fue el haber hallado un modo de financiar el programa, cubriendo los costos con el cobro de la inscripción. Como se verá, la ampliación del programa generó su propia financiación.

Descripción del sistema de selección

El primer examen nacional a los aspirantes a la educación superior se hizo en 1968. En ese año, el Servicio fue adscrito al Instituto Colombiano de Fomento de la Educación Superior (ICFES), heredero del Fondo, entidad oficial, con autonomía administrativa adscrita al Ministerio de Educación.

El programa principal del Servicio Nacional de Pruebas ha sido hasta 1990 la realización de exámenes para el ingreso a la educación superior. Otros de sus programas son la validación del bachillerato académico o de grados del mismo para estudiantes adultos o estudiantes que cursan programas por radio; exámenes de ingreso a la carrera diplomática; exámenes para admisión a grados superiores de la carrera de medicina y expedición de certificaciones (Cuadro 1).

Los exámenes pasaron por dos fases hasta 1990:

1. Entre 1968 y 1980 eran voluntarios para las universidades que quisieran contar con este instrumento de selección. Pero los examinandos crecieron de forma continua, pasando de 26.253 en 1968 a 108.268 en 1980 (Cuadro 1). En 1975 el 39 por ciento de las instituciones de educación superior solicitaban datos sobre los aspirantes.
2. A partir de 1980, el gobierno estableció la obligatoriedad de los exámenes. Sin embargo, para conservar la autonomía, los puntajes podían ser una condición variable entre varios requisitos que cada universidad ponderara. La cobertura del sistema de selección casi se duplicó al pasar de 150.267 en 1980 a 275.152 en 1989 (Cuadro 1).

La creciente demanda de exámenes, unida a los otros programas conexos, hizo posible una creciente sistematización del servicio, un aumento de la capacidad técnica y, sobre todo, una oferta de múltiples servicios que en conjunto dejan un superávit considerable cada año, como se verá más adelante.

La estructura de la prueba obedece a la taxonomía de los objetivos de la educación establecidos por Benjamín Bloom (Bloom y otros, 1973; Acero, 1990: 98-103). De los tres objetivos, la prueba examina el desarrollo cognoscitivo, dejando de lado el afectivo y el psicomotor. El desarrollo cognoscitivo se divide a su vez en funciones de memoria, comprensión, aplicación, análisis y evaluación (Bloom, 1973),

funciones que determinan la estructura de las preguntas.

Cuadro 1. Número de personas que se presentaron al Examen Nacional

Año	Total de alumnos examinados
1968	26.253
1969	32.253
1970	44.339
1971	50.747
1972	51.650
1973	53.498
1974	55.662
1975	75.907
1976	80.337
1977	94.689
1978	95.757
1979	108.268
1980	150.267
1981	174.397
1982	224.335
1983	223.785
1984	223.938
1985	228.272
1986	240.442
1987	250.104
1988	265.147
1989	275.152
1990	301.073
1991	297.143
1992	306.877
1993	338.534
1994	379.827
1995	436.176
1996	480.611

Fuente: Series estadísticas históricas de los programas del Servicio Nacional de Pruebas. 1968 - 1997. ICFES, SERVICIO NACIONAL DE PRUEBAS. 1990

El Servicio refleja los programas de estudio oficiales que hasta 1994 eran únicos para el país. Las pruebas comprenden una parte común y una electiva. La parte común consiste en un examen de aptitud verbal y otro de aptitud mátemática, a los cuales se añaden 5 pruebas de conocimientos, a saber: biología, química, física, español y literatura y ciencias sociales. La octava prueba es una electiva entre 12 opciones, la mayor parte de las cuales responde a los énfasis derivados del

bachillerato diversificado (últimos dos años).

Las preguntas eran diseñadas hasta 1991 por psicólogos, psicómetras, estadísticos y maestros de educación primaria, secundaria y superior procedentes de distintas partes del país.

El examen comprende cerca de 460 preguntas de selección múltiple, con una respuesta verdadera y cuatro distractores. De ellas, hay 50 que no valen para el puntaje, porque se experimentan, luego de pasar por pruebas previas. Existe un banco con cerca de 35.000 preguntas que están siendo incluidas en el sistema computarizado.

Tras la lectura óptica de las hojas de respuesta, el proceso de calificación es realizado en forma electrónica (Caro, 13). El número de respuestas correctas constituye el puntaje bruto, el cual es convertido a una escala estándar, cuyo puntaje para cada una de las ocho series de preguntas va de 20 a 80, con una media de 50 y una desviación típica de 10. De este modo, se configura una curva normal que compara los resultados diferenciales de los examinados, curva que oscila entre cero y un poco más de 400 puntos. Los resultados pueden agruparse y compararse por áreas, por colegios, por regiones y por diversas cohortes a través del tiempo. Todo el proceso está sistematizado.

Las pruebas se administran en día y medio, con tres bloques de tres horas cada uno, hacia abril o mayo para colegios de calendario B (que inicia en agosto), y hacia septiembre para los de calendario A (que comienza en febrero) que son la mayoría. En ambos casos, pueden presentarse estudiantes que lo hayan hecho en ocasiones anteriores. Las pruebas se ofrecían hasta 1990 en cerca de 148 municipios , 488 edificios y 8.800 salones (Caro, 10), según el banco de instituciones de educación secundaria y educación superior que mantiene actualizado el sistema del servicio. Los coordinadores y jurados se escogen por concurso entre maestros y estudiantes universitarios.

Las citaciones se divulgan por distintos medios, entre ellos la prensa nacional y local. El Servicio Nacional de Pruebas incluye en su base de datos los establecimientos de educación secundaria y a ellos se dirigen los formularios de inscripción, por medio de los directores.

El Servicio Nacional de Pruebas cuenta con la imprenta del ICFES y con sistemas modernos de cómputo. Los resultados se entregan al mes de realizado el examen, por medio de los directores de los colegios. Los estudiantes pueden solicitar también los certificados por correo o acudiendo a una central telefónica. El Servicio publica periódicamente estadísticas de población (características de los egresados, orientación hacia la educación superior); informes sobre el nivel académico de los planteles de educación media (divididos en rendimientos altos, medios y bajos) y listados de los estudiantes con puntajes más altos por municipio.

Especial cuidado presta el Servicio Nacional de Pruebas al secreto en el diseño, y la prueba y guarda de las preguntas que elaboran los equipos coordinados por el Servicio. Lo mismo ocurre con la impresión, el transporte, la recolección y el procesamiento de resultados que, salvo este último, se confían a una empresa de transporte de valores. Sólo en dos ocasiones se ha presentado riesgo de fraude, por robo de cuadernillos, pero en cada caso se detectó a tiempo.

Impactos del sistema de selección

Ha existido mucha confianza sobre los resultados de las pruebas, aunque en ocasiones ha habido críticas, sobre todo por la sobrevaloración de la dimensión cognoscitiva respecto a otros componentes. A este respecto, el Servicio insiste en reiterar que las pruebas no se han diseñado para evaluar la calidad de la educación media, aunque puedan ser una fuente indirecta de ella y de hecho muy valiosa.

Se pueden indicar siete medidas de impacto de los exámenes de selección:

1. El uso de los resultados como mecanismo de selección. De un total de 212 universidades existentes en 1989, veintiocho (13,2 por ciento) usaban el puntaje del examen del Servicio Nacional de Pruebas como criterio único, veintinueve (13,7 por ciento) añadían pruebas escritas aplicadas por la universidad[2], ochenta y tres universidades (39,2 por ciento) integraban los resultados del examen del Servicio con entrevistas personales, cincuenta y seis (26,4 por ciento) lo cruzaban con pruebas universitarias y con entrevistas, dos (0,8 por ciento) combinaban las pruebas del Servicio con pruebas universitarias y notas de bachillerato, cinco (2,35 por ciento) ponderaban el puntaje del Servicio con entrevistas y notas de bachillerato, cuatro (1,9 por ciento) añadían al examen pruebas universitarias, entrevistas y notas de bachillerato y cinco universidades (2,35 por ciento) complementaban el examen con notas de bachillerato (Benavides, 1989: 100-101).

2. La cobertura en los medios de comunicación. Tanto la convocatoria de los exámenes, como los resultados de los mismos, son difundidos de modo amplio por la prensa nacional y local: se trata de decidir el futuro de cerca de 275.000 hijos de familia en 1990, ó de medio millón, en 1997 y de calificar en forma indirecta a cerca de 5.000 establecimientos de educación secundaria, y por tanto a sus directivos y maestros. En muchas ocasiones la información de prensa y radio se enriquece con debates en los que participan políticos, técnicos, maestros, directivos docentes y padres de familia.

3. Incidencia directa de los exámenes en las familias, los estudiantes y los maestros. Los establecimientos se esfuerzan al máximo por obtener los mejores puntajes promedio, pues de ellos dependen el prestigio, las inscripciones y los costos de las mismas. Los padres de familia de estratos medios o altos de las ciudades grandes toman en cuenta los puntajes promedio del colegio. Todos los estudiantes saben que su oportunidad de ingresar en una buena universidad pública o privada dependerá en buena medida de su puntaje. Los maestros conocen la estructura de la prueba.

La presión por obtener buenos resultados individuales e institucionales es tal que en ocasiones los colegios destinan mucho tiempo del último año a la preparación del examen. En casos extremos, llegan incluso a expulsar a los estudiantes que puedan perjudicar el promedio.

Una consecuencia de la valoración de los exámenes es la existencia de muchísimos centros de educación no formal en las ciudades, los cuales prometen preparar con éxito a los estudiantes. En un caso excepcional, un centro fue el origen de una universidad. Unos pocos ofrecen servicios técnicos de orientación, según las entrevistas. En su mayoría son, sin embargo, un negocio fácil, que impone una instrucción memorística y mecánica.

4. Creación de un sistema de recompensas. La Empresa Colombiana de Petróleos premia a los mejores bachilleres con la financiación de los estudios universitarios. El Ministerio de Educación instauró en 1982 la orden Andrés Bello que otorga un reconocimiento a los cien mejores bachilleres. La Universidad Nacional atrae a los mejores bachilleres con la entrega gratuita de los formularios de inscripción. Los mejores puntajes enriquecen la hoja de vida y son criterio importante para conceder créditos educativos (El Universal 1996, noviembre 4: 7 A). La prensa regional y local destaca a los mejores bachilleres y a los mejores colegios.

5. Uso de los exámenes como fuente de información indirecta sobre la calidad de la educación y sobre el bachiller. Los exámenes de estado han sido la más antigua y, hasta 1991, casi la única fuente de información, aunque indirecta, sobre la calidad de la educación, exceptuados los estudios etnográficos y los indicadores de matrícula, retención y repitencia. Así, por ejemplo, en 1984 (Rojas y otros) y en 1988 (ICFES/CENCO) dos centros de investigación diferentes realizaron, con apoyo del Servicio, un perfil del bachiller colombiano, basándose en la información socioeconómica de los estudiantes que acompaña a la inscripción en los exámenes.

6. Emulación regional y municipal inducida por los exámenes. El caso más importante ha ocurrido en Antioquia y en su capital, Medellín, que han mostrado una tendencia acusada de descenso en la calidad de la educación, por la medición indirecta deducida de la comparación temporal de los resultados agregados de las pruebas (Alviar y Polanía, 1993).

En efecto, entre 1981 y 1990 la tendencia nacional indicaba un ligero aumento de la proporción de colegios de rendimiento alto (del 16,81 por ciento al 17,28 por ciento). En cambio, en Medellín ocurrió un descenso notorio (del 24,34 por ciento al 20,31 por ciento). En el mismo período, la tendencia nacional mostraba un aumento agudo de los colegios de rendimiento bajo (del 38,23 por ciento al 6,10 por ciento). En Medellín, sin embargo, la tendencia fue mucho más pronunciada (del 36,84 por ciento al 57,42 por ciento) (Alviar: 33).

La interpretación de los resultados se convirtió en un auténtico acertijo y suscitó

una controversia nacional y regional muy aguda (El Mundo, 1991, noviembre 11). Al final, las autoridades educativas de Medellín optaron por diseñar, para las pruebas de 1996, el programa "Medellín Primero en el ICFES", por el cual organizaban un curso oficial preuniversitario masivo, reconocían becas al millar de bachilleres con los mejores puntajes para estudiar en universidades públicas regionales y concedían estímulos a los mejores colegios (Prensa antioqueña, 1996, octubre). Pero, además, con visión estratégica, el Departamento de Antioquia gestionó un crédito externo por 40 millones de dólares americanos para mejorar la educación departamental (Marulanda, 1997: 2A).

7. Capacidad del Servicio para solucionar situaciones de crisis. En una ocasión el Servicio conjuró la intención de un Ministro de Educación que expresó su deseo de acabar con el examen de estado (prensa nacional y regional, 1988: 15 de julio). Luego, el Servicio anuló iniciativas legislativas tendentes a suprimir o a modificar en forma radical los exámenes (Cámara de Representantes, 1992). Más adelante, hubo de apelar al Pleno de la Corte Constitucional para anular una disposición legal, sancionada por el presidente, que concedía un 10% más de puntaje a los alumnos que prestaran el servicio militar (Congreso de la República, 1993; Corte Constitucional, 1996).

Balance de éxitos y limitaciones

Dos razones explican el éxito del Servicio Nacional de Pruebas. La primera, un liderazgo carismático y técnico ejercido con una continuidad excepcional: desde su fundación hasta 1989, el Servicio contó con sólo dos directores. Desde entonces al presente ha habido otros cuatro que, pese a no pocos cambios, han respetado y enriquecido el patrimonio técnico.

La segunda razón consiste en que el Servicio se financia con sus distintos programas. "El costo real total de los 300.637 exámenes de 1990 fue de 688.5 millones de pesos (de tal año), siendo el costo unitario de 2.290 pesos" (Caro, 16). Dado que el cambio en diciembre 19 de 1990 era de 562,29 pesos por dólar americano, las cifras equivalen a 1,22 millones y a 4,07 de dólares. Como entonces se cobraba una tarifa de 1.500 pesos, equivalente a 2,66 dólares americanos., otros programas del mismo Servicio debieron cubrir el subsidio. No obstante, para 1991 la tarifa se duplicó.

Hasta 1990, dos fueron las limitaciones principales. Como ocurre en América Latina con los sistemas de evaluación (Horn y otros, s.f.), la información producida ha sido subutilizada, pese a que su uso real no sea desdeñable. Esto se explica porque hasta 1990 hubo un déficit de investigación en el Servicio y fuera de él.

Más grave ha sido la segunda limitación, que es de orden institucional y que hasta el momento no ha sido superada: el Servicio Nacional de Pruebas ha estado "atado" al Instituto Colombiano de Fomento de la Educación Superior, unas veces como División, otras como unidad especial, y otras como Subdirección. Tal relación ha

sido negativa en más de un sentido : ha limitado el sistema a los exámenes de selección y a otras variantes afines, así como su capacidad administrativa para liderar un sistema nacional de evaluación.

En distintas ocasiones se pensó hacer del Servicio Nacional de Pruebas un instituto autónomo, que aunara las funciones de selección universitaria, evaluación de la calidad, investigación e incluso formación de recursos humanos en evaluación para la educación en general, pero halló la resistencia del ICFES. El Servicio financia sus operaciones con los recaudos por sus distintos programas y genera un superávit considerable, como se verá más adelante, mientras que el ICFES depende para la mayoría de las operaciones de control y fomento del presupuesto asignado por el gobierno.

Todo ello ha dificultado el paso de funciones de selección universitaria a funciones de evaluación de la calidad de la educación. Este paso de un sistema de selección a otro de evaluación es complejo, porque se trata de sistemas que responden idealmente a lógicas diferentes (Greaney y Kellaghan, 1996: 31-39). Así, y teniendo presente el caso colombiano, podemos decir que:

1. Uno es un sistema de selección y de asignación de futuros papeles y de recompensas asociadas al éxito, mientras que el otro es un sistema de reasignación de recursos financieros, físicos y humanos en función de la equidad y del logro aún no alcanzado, pero posible. Por tanto, el primero favorece el llamado "efecto Mateo" (Merton, 1973) , es decir, la concentración de oportunidades en los que más tienen, mientras que el segundo está diseñado para contrarrestarlo. El primero predice un comportamiento futuro, el segundo anticipa un comportamiento susceptible de mejorar y señala cómo y dónde hacerlo.

2. Uno se guía por la norma, mientras que el otro lo hace por el criterio. Por tanto, el primero discrimina, mientras que el segundo diferencia.

3. Uno tiende a situarse en la fase terminal de un gran ciclo (13 años, incluido preescolar) e inicial de otra secuencia (10 años, si se incluye el doctorado), el otro sigue distintos cortes en el tiempo.

4. Uno tiende a concentrarse en contenidos de saber, abstraídos de los demás, mientras que el otro se obliga a integrar desde el comienzo múltiples perspectivas (por ejemplo, factores asociados al logro). Uno tiende a condensar, desplazar e invertir (en el sentido psicoanalítico) todo el saber que se considera válido (Díaz Barriga, 1993), mientras que el otro puede aproximarse al saber desde dimensiones parciales, para integrarlas cada vez más. Uno tiende a manipular la ignorancia enciclopédica, mientras que el otro tiende a reconocer dimensiones más modestas, pero múltiples en indagaciones.

5. Uno tiende a ser decidido desde una posición de autoridad en el saber, el otro presupone saberes ubicuos y en construcción.

6. Uno tiene una realimentación escasa porque su fuente de información, aunque sea universal (el conjunto de estudiantes de 11 grado), es única (los estudiantes), otro la aumenta por multiplicar los evaluadores (autoevaluación y evaluación externa), las fuentes, los métodos de indagación (cuantitativos y cualitativos) y los usuarios (padres de familia, autoridades locales, maestros, investigadores).

7. Uno tiende al secreto y a la reserva, el otro a la comunicación y divulgación.

Pero esta oposición puede ser más dúctil que la descrita, sobre todo cuando los dos sistemas se complementan, como sucedió a partir de 1990.

SISTEMA NACIONAL DE EVALUACIÓN DE LA CALIDAD: 1991-1997

El marco de la educación: 1991-1997

Desde que se comenzara a percibir en 1976, la crisis de la educación suscitó un interés de reforma, especialmente en el seno del gobierno y del magisterio escolar:

1. La preocupación del gobierno central porque un mayor gasto expandía coberturas, pero la repitencia y la deserción anulaban los logros. Para remediarlo, el gobierno empleó distintas estrategias entre 1976 y 1989. Éstas fueron una forma de control racional—siempre limitado— impuesto por capas modernizadoras que poseían autoridad técnica, pero poco poder político: el diseño de un mapa educativo, conformado por distritos y núcleos para favorecer la supervisión; la renovación curricular; la introducción de centros experimentales piloto; la vigilancia de las relaciones intergubernamentales; la aprobación del Estatuto Docente y el arreglo de las prestaciones sociales del magisterio; la flexibilización curricular y la promoción automática; las reformas al Ministerio de Educación; y la estrategia de la Escuela Nueva para la educación primaria rural.

2. El movimiento pedagógico conformado por maestros que se propusieron asimilar fuentes heterogéneas para reformular el papel del maestro y de la educación (Caviedes, 1975; Martínez y otros, 1994; Medina, 1996). No pocos estudios e investigaciones apelaron a teorías y métodos marxistas, estructuralistas, psicoanalistas, etnográficos, hermenéuticos, lingüísticos, neopositivistas, neomodernistas, constructivistas y postreproduccionistas. Ellos enriquecieron el pensamiento sobre la educación (Díaz Villa, 1993). Los estudios etnográficos sobre la escuela y el joven revelaron la profundidad de la crisis (Parra, 1996; Pérez y Mejía, 1996).

Estos intereses, en apariencia disímiles, convergieron, sin embargo, con muchos otros factores, en un nuevo pacto social, que se plasmó en la Constitución de 1991. Ésta proclamó un estado social de derecho con amplias libertades; reconoció la pluralidad cultural; amplió el poder ciudadano en la elección directa de las

autoridades locales y regionales y en la vigilancia y el control indirectos de la gestión pública; y consagró la progresiva descentralización de muchos servicios básicos, cuya prestación trasladó a los departamentos y municipios, junto con la cesión de una porción creciente de ingresos nacionales.

Estos cambios inspiraron la Ley General de Educación (ley 15 de 1994), en la elaboración de la cual participó el magisterio. La ley descentralizó las responsabilidades curriculares para otorgar autonomía a directivos, docentes, maestros, alumnos y padres de familia en la elaboración de proyectos educativos institucionales. Con la intención de mantener parámetros nacionales, la ley se propuso fortalecer el Ministerio de Educación Nacional, el cual debía fijar un plan decenal concertado, instaurar un sistema nacional de información, señalar áreas mínimas e indicadores de logro y organizar el sistema de evaluación. Este debe conciliar la autoevaluación y la evaluación externa, integrando en esta última la evaluación de maestros, pedagogías, proyectos educativos institucionales, textos y materiales.

Antecedentes del sistema de evaluación

En los últimos quince años, el factor que más ha incidido en despertar la preocupación por la calidad de la educación en Colombia y en estimular un sistema para evaluarla ha sido el programa de Escuela Nueva. Ésta ha sido una innovación colombiana de alcance internacional, pues es un modelo que soluciona con gran imaginación problemas de cantidad, calidad y eficiencia. Constituye quizás la mejor alternativa dentro de la educación formal para ofrecer una educación primaria completa de calidad (cinco años) a los sectores más pobres y dispersos del campo y aun de la ciudad.

Es imposible hacer un resumen del programa sin empobrecer su significado, por lo cual lo mejor es remitir a una bibliografía mínima sobre su desarrollo (Colbert y otros, 1976; Torres, 1996; Schiefelbein y otros, 1996; Psacharopoulos y otros, 1996). Es un sistema multígrado y modular, intensivo en textos para alumnos y guías para el maestro, participativo, inspirado en la escuela activa, abierto a la comunidad y a la ecología.

Su antecedente fue la Escuela Unitaria Completa, promovida desde 1960 por el Programa Principal de la UNESCO y derivada de muchas experiencias pedagógicas del siglo XX (Hernández, 1961). En Colombia, se había iniciado en una pequeña ciudad, Pamplona, hacia 1962. A mitad de la década, había 150 escuelas. En 1967, el gobierno extendió el modelo a todas las escuelas con un solo maestro (Torres, 1996). El perfil actual como Escuela Nueva fue definido entre 1975 y 1978, cuando llegaron a funcionar 500 escuelas en tres departamentos.

Los éxitos de la Escuela Nueva atrajeron el interés del Banco Mundial, el cual financiaría su expansión en dos fases: la primera, de consolidación regional y de experimentación, entre 1982 y 1988, y la siguiente de universalización, de 1989 a

1997 (Duarte, 1995: 126-147). En 1985 ya había 8.000 escuelas, en 1989 eran 17.984 y en 1991 llegaban a 20.000 de las 27.000 escuelas rurales del país (Torres,1996).

El interés del Banco Mundial era comprensible, pues había aceptado abrir líneas de crédito a la educación en el mismo año, 1979, en que la Academia de Suecia concedió el nóbel de economía a pioneros de la teoría del capital humano (Duarte, 1995: 127).

Los institutos privados de investigación de muy buena calidad, los organismos multilaterales y bilaterales y el gobierno alternaron esfuerzos desde 1980 por medir la calidad de la enseñanza de la educación primaria y comprender el significado de la Escuela Nueva. El Instituto SER[3], con apoyo del Centro Internacional de Investigaciones para el Desarrollo del Canadá, diseñó una evaluación pionera de logro para estudiantes de tercer y quinto grados en las áreas de lenguaje y matemáticas (Rodríguez, 1982). Dicha investigación fue el primer paso en el establecimiento de un sistema nacional de evaluación. Es de notar que el investigador, ahora consultor de una institución privada, había sido el fundador mismo del Servicio Nacional de Pruebas.

La investigación, realizada por muestra, estrenó la metodología que desde entonces se ha seguido para el análisis de logro. El autor distinguió entonces, con claridad, la diferencia entre pruebas de norma, como las realizadas para la selección a la educación superior y pruebas según criterio, como las contenidas en la investigación sobre logro en lenguaje y matemáticas (Rodríguez, 1982: 35-38).

Una de las conclusiones de la investigación era decisiva: "en general, los porcentajes de logro en la Escuela Nueva, que comprende básica y esencialmente escuelas rurales, son significativamente superiores a los de sus homólogas de escuela tradicional, las de zonas rurales, y aun a las de las localidades urbanas menores" (pág. 267). Dicha conclusión es demostrativa de la excelencia pedagógica de la Escuela Nueva, pues ésta operaba en zonas de mayor carencia.

El siguiente paso fue elaborar una metodología para explicar los factores asociados al logro y contrastar de modo más específico las diferencias entre la Escuela Nueva y la escuela tradicional. Esto fue realizado por el Instituto SER, a petición del Ministerio de Educación. Uno de los investigadores había ordenado algunas evaluaciones hechas por los propios docentes de la Escuela Nueva, que había resumido con cierto escepticismo sobre el modelo (Rojas y Briceño, 1982).

No obstante, el mismo investigador dirigió un lustro, después una indagación más ambiciosa, de nuevo solicitada por el Ministerio de Educación (Rojas y Zoraida, 1987). Utilizando la misma metodología de criterio para el análisis de logro en matemáticas y lenguaje establecida por Rodríguez, el autor fue más allá, organizando cuestionarios a maestros, directivos, alumnos y padres de familia sobre factores asociados, tales como la autoestima, la creatividad, el comportamiento cívico, la actitud social, la percepción sobre marginamiento y la frecuencia de

exposición a medios de comunicación y a libros, entre otros.

Las conclusiones mostraban de nuevo indudables ventajas de la Escuela Nueva: " a nivel nacional, obtuvo puntajes significativamente superiores a los calculados para la escuela rural graduada en las pruebas de comportamiento cívico social, autoconcepto social, matemáticas de tercero, así como en español de tercero y quinto de primaria" (pág. 191). Además, se estimaba que el costo era apenas 10 por ciento superior a la escuela tradicional, siendo más intensiva la inversión en insumos como textos, materiales y dotaciones.

Tales evaluaciones debieron ser demostrativas para que el Banco Mundial y el Ministerio de Educación decidieran avanzar en el programa de universalización de la educación primaria con el modelo de la Escuela Nueva desde 1988 y, además, para que el Ministerio de Educación decidiera crear, por la Ley 24 del mismo año, una División de Control de la Calidad de la Educación. Esta división propuso, en 1989, un sistema nacional de evaluación con base en las recomendaciones hechas por un grupo consultivo formado por representantes de entidades públicas (Ministerio de Educación, Universidad Nacional y Universidad Pedagógica) y privadas (Instituto SER de investigación).

En 1990 se llevó a cabo otra evaluación de logro con carácter experimental, reducida a tres departamentos, en los cuales el programa de Escuela Nueva había demostrado avances notables. Esta evaluación se llevó a cabo con la participación de maestros y se aplicó a los grados tercero y quinto de primaria, en lenguaje y matemáticas, más los factores asociados.

Entre las instituciones anteriores no estaba presente el Servicio Nacional de Pruebas, pese a que contaba con experiencia en validación de bachillerato, exámenes a estudiantes en el paso de quinto a sexto grado y la evaluación de maestros. Ello reflejaba la diferencia y la tensión entre un sistema de selección y otro de evaluación.

Conformación y ampliación del sistema

En 1991, el Servicio Nacional de Pruebas se incorporó al Sistema Nacional de Evaluación de la Calidad, por invitación del Ministerio. Esto contribuyó a la extensión de la cobertura. El Servicio elaboró pruebas de logro en matemáticas y lenguaje para quinto curso—que fueron aplicadas por el Instituto SER. Éste aplicó a los alumnos de tercero de primaria pruebas en iguales áreas, diseñadas por especialistas contratados. Se añadieron cuestionarios para estudiar factores asociados.

La medición de logro se hizo en 13 entidades territoriales de las 33 que integran el país, con una muestra de 15.000 alumnos de 218 escuelas urbanas y 212 rurales. Luego se amplió a 6.000 niños en los mismos grados y áreas (Ministerio de Educación, 1993). En octubre de 1992, se aplicaron pruebas en las mismas áreas a

25.189 estudiantes de los grados séptimo y noveno en 22 entidades territoriales, más cuestionarios sobre factores asociados. En 1993 se aplicó a 11.591 en cuatro departamentos.

Las discusiones llevadas a cabo en noviembre de 1993 dejaron entrever la complejidad del sistema, por la diferencia de perspectivas sobre la concepción y el manejo de información, la administración, la evaluación, la investigación, la participación regional, la divulgación y la toma de decisiones.

Entre 1993 y 1994 el Ministerio de Educación, el Servicio Nacional de Pruebas y el Instituto SER aumentaron la escala al establecer una línea de base con una muestra nacional diseñada con criterios definidos de común acuerdo, que serviría en el futuro para aplicaciones en áreas no cubiertas, como ciencias naturales y sociales y para repeticiones cuatrienales en cada una de las áreas para registrar cambios temporales.

El Servicio se encargó de revisar los instrumentos de las pruebas de matemáticas y lenguaje y de los cuestionarios de factores asociados. El Instituto SER los aplicó en 1993 a estudiantes de los cursos tercero y quinto del calendario A y el Servicio a estudiantes de calendario B, en 1994. La muestra incluyó en total a 53.000 estudiantes de 1.628 planteles de 33 divisiones territoriales del país.

Más adelante, en diciembre de 1996, se vería la necesidad de precisar una muestra maestra nacional, dado que las aplicaciones anteriores se habían hecho sobre muestras diferentes en cada ocasión, con tamaño y cobertura variables. Esta tarea se encargó a una empresa privada especializada, con criterios definidos por las instituciones más involucradas en el sistema: el Ministerio de Educación Nacional, el Servicio Nacional de Pruebas, el Departamento Nacional de Planeación[4] y el Instituto SER.

Entretanto, el Servicio había iniciado un cambio, estimulado por la nueva responsabilidad. Desde 1992, los procesos administrativos se estandarizaron en un detallado manual de funciones, de tal forma que cada fase realimentara de modo preciso a la siguiente (Páez, 1992). Una importante inversión en sistematización ocurrió en 1995. Si hasta entonces las actividades manuales ocupaban cerca de 70 por ciento, ahora son los procesos computarizados los que representan tal porcentaje (Páez, 1997).

Pero éstos no fueron los únicos cambios. Los ahorros en tiempo y en recursos se emplearon en una reestructuración conceptual y en una ampliación de recursos humanos. Hasta 1990, el Servicio había sido deficitario en investigación. A partir de entonces se crearon líneas permanentes de investigación que conjugaban lo cualitativo con lo cuantitativo. Hasta 1991, el Servicio había dependido de la psicología. Sin que esta profesión dejara de ser crucial, se constituyeron equipos especializados en las cuatro áreas fundamentales: lenguaje, matemáticas, ciencias y ciencias sociales. El trabajo interdisciplinario era bastante intenso. Se abrió la

discusión con los maestros, las universidades y las regiones.

El Servicio asumió con eficiencia y creatividad los nuevos retos del sistema nacional de evaluación, sin descuidar sus tareas en el sistema de selección. Desde 1995 la taxonomía de Bloom se ha conciliado con la introducción de cuatro competencias derivadas de la teoría de la acción comunicativa (Habermas, 1987) : competencias comunicativa, científico-técnica, estética y ética (ICFES. SNP. 1997, a, b, c, d, e). A partir de 1994, el Servicio tomó en consideración, para la realización del tradicional examen, los contenidos mínimos y los indicadores de logro fijados como pauta nacional por el Ministerio de Educación. En 1990 se presentaron al examen de admisión 301.073 estudiantes, cifra que ascendió a 480.611 en 1996: ¡casi medio millón de estudiantes (Cuadro 1)!.

El aumento en la cobertura de este programa ha generado una estructura que asegura la solvencia financiera del Servicio Nacional de Pruebas. En 1996, el Servicio Nacional de Pruebas mostraba un superávit de 1.572 millones de pesos, equivalente casi a 1,5 millones de dólares americanos (Cuadro 2).

El Ministerio de Educación y el Servicio Nacional de Pruebas asumieron otro reto mayor: participar en el Third International Mathematics and Science Study (TIMMS), cuyos resultados—entre el cuadragésimo y el cuadragésimo primer lugar—, si bien se podían anticipar, teniendo en cuenta los problemas de la calidad de la educación, pero no eran tan importantes como el ganar una enorme experiencia en comparación internacional de los currículos—que coinciden en un 80 por ciento con patrones internacionales—, aumentar la destreza técnica y administrativa del sistema de evaluación y disponer de una medida de comparación internacional, nacional y regional que se sumara a las anteriores y cuya interpretación ofrecería la oportunidad para introducir mejoras en la educación (Díaz, 1996).

Por las mismas razones, Colombia se ha sumado desde 1995 a los países que realizan un estudio internacional sobre educación cívica, orientado por la International Association for the Evaluation of Educational Achievement. Esta participación coincide con una investigación patrocinada por el Servicio de Pruebas sobre la enseñanza de las ciencias sociales, investigación que utilizará métodos cuantitativos y cualitativos. Con ello, el Servicio extenderá las pruebas de logro a dominios diferentes, a las matemáticas y el lenguaje. En el presente año, el Ministerio de Educación y el Servicio realizarán, además, evaluaciones a los maestros de educación primaria. Y, para el próximo año, se proyectan pruebas en ciencias naturales y sociales.

Impacto

El mayor impacto del sistema nacional de evaluación ha consistido en anticipar y justificar la decisión de extender a todo el país el programa de Escuela Nueva. Como se ha indicado, entre 1985 y 1991 se pasó de una cobertura de 8.000 escuelas

a otra de 20.000. Pero, además, el sistema nacional de evaluación ha generado una metodología para apreciar la diferencia en el logro y en los factores asociados. Tal metodología podrá extenderse a la ponderación de distintas estrategias pedagógicas, tanto públicas, como privadas, no sólo en la tradicional educación básica primaria, sino además en la secundaria.

Cuadro 2. Servicio Nacional de Pruebas: recaudos y costos de los programas en 1995. Dólares E.E.U.U.

** El déficit en esta aplicación fue cancelado por el Ministerio de Relaciones Exteriores durante la vigencia de 1996. La tasa de cambio del dólar de E.E.U.U. era de 1.000 pesos por dólar.*
*** Costos incluidos dentro de los de cada programa.*

Lo anterior converge con progresos extraordinarios en los indicadores sociales y en las teorías sobre la incidencia del gasto público en Colombia, en general, y, en particular, sobre la importancia del gasto en educación. Un estudio conjunto del Banco Mundial y de la Misión Social del gobierno colombiano ha hecho un diagnóstico preciso sobre la magnitud de la pobreza (May, 1996). Algunos estudios paralelos (Londoño, 1995; Vélez, 1996), han examinado con precisión la relación entre distribución de ingresos, gasto público y desarrollo económico y social. Los

planificadores del gobierno, investigadores de instituciones macroeconómicas y sociales y benefactores de fundaciones sociales usan cada vez más los indicadores de calidad. Una conclusión se impone cada vez más: una inversión bien orientada en educación contribuye al mismo tiempo a la superación de la pobreza y al crecimiento económico. A esta conclusión apuntan las distintas publicaciones de la revista Coyuntura Social , publicación bianual editada desde diciembre de 1989 por las dos instituciones más prestigiosas de investigación económica y social: FEDESARROLLO[5] y el Instituto SER de investigación. Esta publicación surgió en buena medida inspirada por los avances en el sistema de evaluación.

Limitaciones

¿Por qué entonces no se han utilizado los resultados del sistema nacional de evaluación de la calidad de la educación para las políticas y la programación social? O, en otros términos, ¿por qué pese a los avances del sistema nacional de evaluación, su uso no es aún más restringido?

Existe, en primer lugar, un problema de coordinación, que sin duda es notorio, entre la política social y macroeconómica y el sistema nacional de evaluación. El Sistema Nacional de Evaluación está regido por un híbrido: el Ministerio de Educación fija las directrices y destina los recursos específicos para la evaluación (que son diferentes a los recursos del sistema de selección), pero la capacidad técnica y operativa la posee el Servicio Nacional de Pruebas, el cual está supeditado a otra Institución, el ICFES, para el cual no son apreciables las actividades del Servicio, aunque sea envidiable su suficiencia financiera. No obstante, la solución a esta diferencia ya está a la vista, comoquiera que el actual Ministro de Educación ha enunciado el propósito de establecer un instituto de medición de la calidad de la educación, con autonomías financiera, técnica y operativa.

Con todo, la segunda dificultad es más apreciable y requerirá de mayor dedicación política y técnica: se trata de una deficiencia notoria en el modo como se realizó la descentralización de los servicios sociales y, en particular, la de la educación. El defecto principal consiste en que la gestión territorial de los recursos trasladados no está unida de modo orgánico a la gestión de la educación, porque la ley no ofreció a la escuela o al colegio instrumentos financieros o administrativos para ejercer la autonomía (Duarte, 1997; Sarmiento y Vargas, 1997). En otros términos y a diferencia de otros países, la descentralización no ha sido diseñada para movilizar recursos financieros y humanos en función de una combinación óptima de necesidades y de calidad de los servicios. Se presenta una situación aberrante según la cual "mientras más pobre sea el municipio, menos docentes le asigna el departamento" (Sarmiento y Vargas, 1997). Con ello se limita el poder de un sistema nacional de evaluación de la calidad de la educación como indicador de la orientación de recursos.

CONCLUSIONES

Este estudio ha presentado la diferencia que existe, según los expertos internacionales, entre un sistema nacional de selección y uno nacional de evaluación. Es posible, como se ha visto, que un sistema abra paso al otro pero, por tratarse de lógicas diferentes, sería quizás más apropiado establecer desde el principio un sistema de evaluación, del cual la selección fuera apenas una parte.

No obstante, la relación entre uno y otro depende de los contextos históricos e institucionales. En el caso de Colombia, la ventaja de haber partido del sistema de selección ha consistido en primer lugar en ganar una experiencia invaluable en competencia técnica y organizativa, lo cual incluye la administración de un sistema computarizado con alta eficiencia. Este sistema alcanzó en 1996 un impacto considerable, comoquiera que aplicaba pruebas de conocimiento para el paso a la educación superior a cerca de medio millón de estudiantes.

La segunda ventaja, no menos importante, radica en el hecho de que el sistema de selección ha consolidado una autosuficiencia financiera, que siempre será requisito para una continuidad estratégica y para una cierta independencia técnica. Dado que existe una gran elasticidad determinada por el crecimiento potencial de la educación superior, y dado que los costos de los distintos exámenes de selección pueden ser pagados por una población para la cual el acceso a la educación superior significa por lo general un subsidio extraordinario, muchos países de la región podrían optar por esta vía, teniendo el cuidado de desarrollar poco a poco una capacidad paralela para la evaluación de la educación primaria y secundaria. Lo más esencial en este punto consiste en escoger el problema más pertinente, como fue el caso de la Escuela Nueva en Colombia.

Asunto decisivo en cualquier caso será la definición institucional del Sistema Nacional de Evaluación de la Calidad. Lo óptimo sería lograr lo que quizás sea una meta próxima de Colombia: crear un instituto independiente—ojalá mixto—que cuente con la mayor competencia técnica y la mayor credibilidad social. En el caso de Colombia, la alianza establecida durante algún tiempo con el Instituto SER de investigación probó ser muy estimulante.

Finalmente, destaquemos la importtancia que tiene la capacidad de interlocución entre las políticas sociales, en particular las políticas de educación y las macroeconómicas. El medio obligado para ello son los sistemas de información y los de evaluación. Políticas bien diseñadas en este sentido, profesional y técnicamente competentes, pueden ser una vía de transacción más efectiva que muchos discursos sobre la necesidad del gasto social.

Este diálogo depende en buena medida del modo como haya sido concebida la descentralización en un país y en particular de la forma en que el balance de la educación distribuya responsabilidades entre el gobierno central, los departamentos y los municipios. Hay en ello una dimensión política que trasciende la dimensión

técnica, pero es indudable que ésta, bien orientada, puede incidir en cambios políticos.

NOTAS

[1] El autor desea agradecer en particular a las siguientes personas, entre otras muchas, que contribuyeron con amabilidad a la realización de este ensayo: Benjamín Álvarez (AED), Magdalena Mantilla (Jefe de la Subdirección Servicio Nacional de Pruebas del Instituto Colombiano de Fomento de la Educación Superior); Fernando Páez, del Servicio Nacional de Pruebas (SNP), una Subdirección del Instituto Colombiano de Educación Superior (ICFES). Dependencia adscrita al Ministerio de Educación Nacional de Colombia. Funcionarios del Servicio Nacional de Pruebas; Blanca Otálora (Directora de Organización escolar del Ministerio de Educación Nacional); Jesús Duarte (Jefe de la Unidad de Desarrollo Social del Departamento Nacional de Planeación); Alfredo Sarmiento (Jefe de la Misión Social del Departamento Nacional de Planeación); Pedro Amaya, director del Instituto SER de investigación; Carmenza Bulla por la realización de entrevistas. Maestros, directivos docentes y profesores universitarios. No obstante, las afirmaciones contenidas en este documento son de responsabilidad entera del autor y no expresan necesariamente el pensamiento de las instituciones o de las personas consultadas.

[2] La Universidad Nacional se regía por su propio examen de admisión, similar al del Servicio.

[3] Instituto SER de Investigación. Entidad privada e independiente de investigación, sin ánimo de lucro, creada en 1974, con vocación para el estudio de sistemas de las organizaciones y, en especial, de la investigación pionera sobre temas de justicia, seguridad ciudadana, seguridad social, educación, salud, transporte y gestión pública. Cuenta con más de 300 informes sobre sectores estratégicos.

[4] DNP. Departamento Nacional de Planeación. Sus antecedentes datan de 1950. En la actualidad es un Departamento Administrativo adscrito a la Presidencia de la República, el cual ejerce la secretaría del Consejo Nacional de Política Económica y Social, encabezado por el Presidente y encargado de determinar los planes de inversión, sujetos al plan general de desarrollo que debe ser aprobado por el Congreso y que debe sufrir un proceso de concertación nacional.

[5] FEDESARROLLO. Entidad privada e independiente de investigación, sin ánimo de lucro, creada en 1970 y especializada en política macroeconómica. Además de muchos libros e informes económicos, edita la revista Coyuntura Económica y, desde 1989, con el Instituto SER de investigación, la revista Coyuntura Social . Su historia está relatada en el excelente libro de Gómez Buendía, Hernando (editor): Economía y Opinón . Santafé de Bogotá: Tercer Mundo.

[6] ICETEX. Instituto Colombiano de Estudios Técnicos en el Exterior. Entidad adscrita al Ministerio de Educación Nacional. Fue creado en 1950 e inició operaciones en 1953. Su gestor, Gabriel Betancur Mejía, diseñó esta novedosa institución a partir de su propia reflexión como estudiante en los Estados Unidos, hacia 1944, habiendo sido usufructuario de un crédito inusual de una empresa antioqueña.

REFERENCIAS

Acero, Hugo. (1990). *Desarrollo histórico de las pruebas de admisión en la universidad colombiana. Servicio Nacional de Pruebas.* Bogotá. Mimeo (ICFES.SNP).

Alexander, Jeffrey 1991. (1988). Sociología cultural: lo sagrado y lo profano en el discurso tecnológico. *Revista Mexicana de Sociología*, abril-junio: 283-309.

Álvarez, Benjamín. (1995). "Equity and selective access to higher education in Latin America". En: Kellaghan, Thomas (Ed.) (1995). *Admission to higher education: issues and practice* . Dublin, Educational Research Centre. Princeton, NJ: International Association for Educational Assessement: 210-222.

Alviar, Mauricio y Polanía, Doris. (1993). "La calidad de la educación." *Educación y Cultura*, *29*, marzo: 24-36. Bogotá.

Ardila, Rubén. (1973). *La Psicología en Colombia. Desarrollo Histórico.* México: Trillas.

Ardila, Rubén. (1993). *Psicología en Colombia. Contexto Social e Histórico*. Bogotá: Tercer Mundo.

Benavides, Álvaro. (1989). *Admisión a la educación superior: panorama internacional.* Bogotá: Universidad Nacional.

Betancur Mejía, Gabriel. (1997). *Entrevista concedida a Gabriel Restrepo sobre el planeamiento integral de la educación, la creación de ICETEX*[6] *y la evolución de la educación*. Bogotá, mecanografiado.

Bloom, Benjamín y otros. (1973). *Taxonomía de los objetivos de la educación. La clasificación de las metas educacionales.* Buenos Aires: El Ateneo Editorial.

Caro, Blanca Lilia. (1990). *Impacto del Servicio Nacional de Pruebas.* Bogotá: Instituto SER, IFT-196, mecanografiado.

Caviedes, Sergio. (1975). *Tecnología educativa y Satélite educativo*. Bogotá: Suramericana.

Colbert de Arboleda, Clara Victoria; Mogollón, Oscar y Levinger, Beryl. (1976). *Hacia la Escuela Nueva: unidades de instrucción y formación para el Maestro Rural que tiene más de un nivel a su cargo*. Bogotá: Ministerio de Educación Nacional y Universidad Pedagógica, mecanografiado.

Congreso de la República. (1993). *Ley 48 por la cual se reglamenta el servicio de reclutamiento y movilización*. Bogotá, mecanografiado.

Coombs, Philip. (1971). *La crisis de la educación mundial.* Barcelona: Península.

Corte Constitucional. (1996). *Demanda inconstitucional contra el artículo 40, literal b, de la Ley 48 de 1993.* Bogotá, mecanografiado .

Departamento Nacional de Planeación. Misión Social. (1997). *Seminario logro del plantel. Notas sueltas.* Bogotá, mecanografiado.

Díaz, Carlos Jairo. (1996). *TIMSS. Informe ejecutivo.* Cali. Mecanografiado.

Díaz Barriga, Angel, compilador. (1993). *El examen: textos para su historia y debate.* México: Universidad Nacional Autónoma de México.

Díaz Villa, Mario. (1993). *El campo intelectual de la educación en Colombia.* Cali: Universidad del Valle.

Duarte, Jesús Hernando. (1995). State education and clientelism in Colombia (the politics of state educational administration and of implementation of educational investment projects in two colombian regions). Tesis presentada para el cumplimiento parcial de los requisitos del doctorado en Ciencia Política, Facultad de Estudios Sociales, Universidad de Oxford.

Duarte, Jesús. (1997a). *Entrevista concedida a Gabriel Restrepo sobre el Sistema Nacional de Evaluación*. Bogotá. Cinta magnética .

Duarte, Jesús. (1997b). *Situaciones críticas del sector educativo colombiano y retos hacia el futuro*. Bogotá. Mecanografiado .

Gómez Buendía, Hernando y Losada Lora, Rodrigo. (1984). *Organización y Conflicto: la educación primaria oficial en Colombia*. Otawa: CIID.

Greagney, Vincent y Kellaghan, Thomas. (1996). *Monitoring the learning outcomes of educational systems*. Washington, DC: Banco Mundial.

Habermas, Jürgen. (1987). Teoría de la acción comunicativa. Madrid: Taurus. Dos tomos.

Horn, Robin, Wolff, Laurence y Vélez, Eduardo. (s.f.) *Developing educational assessment systems in Latin America. A Review of issues and recent experience* . Washington, DC: Banco Mundial.

Kogan, Maurice. (1992). "El ajuste de la evaluación en el marco de la dirección de la educación." *Revista de Educación*, *299*, septiembre a diciembre: 155-168. Madrid.

Kvale, Steinar. (1992). "La evaluación y la descentralización de los conocimientos." En *Revista de Educación*, No. 299, septiembre a diciembre: 119-141. Madrid

Londoño de la Cuesta, Juan Luis. (1995). *Distribución del ingreso y desarrollo económico*. Bogotá: Tercer Mundo.

Martínez B., Alberto, Noguera, Carlos y Castro, Jorge O. (1994). *Currículo y modernización. Cuatro décadas de educación en Colombia*. Bogotá: Foro Nacional por Colombia- Tercer Milenio.

Marulanda, Iván. (1997). "Mataron un niño en el Japón y dos mellizos en Colombia." *El Espectador* (Bogotá), 5 de julio de 1997: 2 A.

May, Ernesto (coordinador general). (1996). *La pobreza en Colombia . Un estudio del Banco Mundial*. Bogotá: Tercer Mundo.

McLuhan, Marshall. (1985). *La Galaxia de Gutemberg.Génesis del Homo typographicus*. Barcelona: Planeta.

Medina Gallego, Carlos. (1996). *Caja de Herramientas para Transformar la Escuela*. Santafé de Bogotá: Rodríguez Quito editores.

Merton , Robert K. (1973). "El efecto Mateo." En: *La Sociología de la Ciencia. Investigaciones teóricas y empíricas*. Madrid: Alianza. Dos tomos.

Nelson, Richard R y Wright, Gavin. (1992). "The Rise and Fall of American Technological Leadership: The Poswar Era in Historical Perspective."*Journal of Economic Literature*, Vol. XXX (December 1992): 1931-1964.

Páez, Fernando. (1992). "Manual de procedimiento para la organización y administración de exámenes en el Servicio Nacional de Pruebas." Bogotá. Tesis de grado presentada a la Escuela Superior de Administración Pública. Mecanografiado.

Páez, Fernando. (1997). *Entrevista sobre la administración del Servicio Nacional de Pruebas*. Bogotá. Cinta magnética .

Parra, Rodrigo. (1996). *Escuela y modernidad en Colombia*. Bogotá: Tercer Mundo. Cuatro tomos.

Pérez, Diego, y Mejía, Marco Raúl. (1996). *De calles, parches, galladas y escuelas . Transformaciones en los procesos de socialización de los jóvenes de hoy*. Bogotá: CINEP.

Psacharopoulos, George, Rojas, Carlos y Vélez, Eduardo. (1996). "Evaluación de resultados en la Escuela Nueva de Colombia. ¿Es el multígrado la respuesta?" *Revista Colombiana de Educación*, *32*: 93-110. Bogotá, Universidad Pedagógica.

Rodríguez, José. (1982). *El uso de los exámenes del Servicio Nacional de Pruebas como indicadores de la calidad de la educación secundaria*. Bogotá: Instituto SER de investigación, IFT-040. Mecanografiado.

Rojas, Carlos y Briceño, Rosa Cecilia. (1982). *Escuela Nueva 1977-1981. Características Generales*. Bogotá: Instituto SER de investigación, IFT-039. Mecanografiado.

Rojas, Carlos y Castillo, Zoraida. (1987). *Evaluación del programa Escuela Nueva en Colombia. Características de los planteles docentes y alumnos; logro de sus estudiantes en comparación con los de las escuelas tradicionales*. Bogotá: Instituto SER de investigación, IFT-133. Mecanografiado.

Rojas, Carlos, Amézquita de Pardo, Helena y González Peña, Germán. (1984). *Características del bachiller colombiano y su relación con los resultados en los exámenes de estado y su ingreso a las instituciones de educación superior*. Bogotá: Instituto SER de investigación, IFT-068. Mecanografiado.

Sarmiento, Alfredo y Vargas, Jorge Enrique. (1997). *Descentralización de los Servicios de Educación y Salud en Colombia: versión preliminar*. Bogotá. Mecanografiado.

Schiefelbein, Ernesto, Vera, Rodrigo, Aranda, Humberto, Vargas, Zoila y Corco, Víctor. (1996). "En busca de la escuela del siglo XXI: ¿puede darnos la pista la Escuela Nueva en Colombia?" *Revista Colombiana de Educación*, No. 32: 20-91. Bogotá: Universidad Pedagógica.

Torres, Rosa María. (1996). "Alternativas dentro de la educación formal: el programa Escuela Nueva en Colombia." *Revista Colombiana de Educación*, 32: 1-19. Bogotá: Universidad Pedagógica.

Weiler, Hans. (1992). "¿Es la descentralización de la dirección educativa un ejercicio contradictorio?" *Revista de Educación*, 299, septiembre-diciembre: 57-80. Madrid.

Parte III

EVALUACIÓN Y PROFESIONALISMO DE LOS MAESTROS

CAPÍTULO 7

EVALUACIÓN DE LOS MAESTROS

Carol Anne Dwyer

La tercera parte del libro se inicia con una revisión crítica del tema de la evaluación de los maestros dentro de la perspectiva general del mejoramiento de la educación, que es el propósito final de la evaluación. Este capítulo presenta un análisis de los propósitos de la evaluación de los docentes, los estándares a los que debe responder y las metodologías disponibles. También identifica aspectos a considerar cuando se ejecutan evaluaciones de los maestros, tales como la selección del método, la inclusión de los resultados de la investigación, los énfasis entre la teoría y la práctica, las diversas perspectivas involucradas y el alcance de los criterios de evaluación.

INTRODUCCIÓN

En este documento se analizarán los aspectos siguientes de la evaluación de los maestros:[1]

- Utilización de la evaluación de los maestros para el mejoramiento educativo en un sistema de evaluación integrado.
- Relaciones entre metas, estándares y pruebas en la evaluación de los maestros.
- Estándares para: validez, calidad técnica y utilización; currículo; conocimiento y aptitud docente; nivel de desempeño; y oportunidad para aprender.
- Fines de la evaluación de los maestros.
- Críticas de las alternativas de evaluación de los maestros.
- Análisis del campo de la enseñanza para fines de evaluación.
- Papel de un concepto rector de la enseñanza para la evaluación de los maestros.
- Selección de los métodos de evaluación.
- Determinadas cuestiones y debates en la evaluación de los maestros.

EVALUACIONES DE LOS MAESTROS Y MEJORAMIENTO EDUCATIVO

Para ser eficaz en el mejoramiento de la educación, la evaluación de los maestros ha de tratarse como elemento intrínseco del sistema educativo. Este es un principio general de la evaluación eficaz y un importante factor en determinar la validez de las evaluaciones utilizadas. Las evaluaciones de los maestros tienen vínculos importantes con los valores sociales del público en general, los de quienes gobiernan y los de los educadores profesionales. Las evaluaciones de los maestros también están vinculadas con las políticas y prácticas de la formación de docentes; con la adquisición de competencia por parte de los maestros que ejercen la profesión; y con las decisiones acerca de los currículos escolares. Es una condición absolutamente necesaria para las evaluaciones de los maestros y los alumnos, en términos de validez, justo equilibrio, carácter práctico y utilidad, hacer claros y coherentes estos vínculos.

Amén de las necesidades técnicas y éticas que contribuyen a satisfacer una visión unificada de la educación, existe una necesidad práctica que ha de satisfacerse también. Tomando como base muchos años de experiencia como diseñadora y redactora de evaluaciones educativas, he llegado a la conclusión de que los datos obtenidos con las evaluaciones que se tratan al margen de los valores de la sociedad y de la vida de las escuelas y las comunidades en las que funcionan, rara vez serán percibidos como útiles por quienes deberían beneficiarse de ellos. A fin de ser beneficiosas para el sistema educativo en su conjunto, estas evaluaciones han de estar cuidadosamente vinculadas con estas entidades mayores. Las características técnicas de las evaluaciones de los maestros y otros tipos de evaluaciones no nos dicen, de por sí, si valdrá la pena realizarlas. El valor de la información derivada de las evaluaciones será determinado, en última instancia, por el grado en el que se vea influenciado el sistema educativo, bien sea positiva o negativamente.

METAS, ESTÁNDARES Y EVALUACIONES

Para comprender el papel que desempeñan las evaluaciones de los maestros en las mejoras educativas se requiere una clara comprensión de las relaciones y distinciones que existen en el sistema educativo en su conjunto. Forman parte integral de este proceso los tres elementos siguientes: metas, estándares y evaluaciones. Estos tres elementos contribuyen de forma importante al mejoramiento de la educación. Si bien las diferencias en su finalidad, aspecto y utilización pueden parecer evidentes, otras no lo son. Es vital establecer vínculos lógicos entre estos elementos y formular políticas y prácticas que concuerden con esta visión más amplia, pero a menudo son pasos que se ignoran al establecer mecanismos eficaces para planificar, impartir y evaluar la instrucción. La Figura 1 proporciona un cuadro simplificado de las interconexiones entre metas, estándares y evaluaciones en el mejoramiento educativo.

Figura 1. Proceso de utilización de las evaluaciones para el mejoramiento educativo

Las metas son inherentemente exposiciones de las aspiraciones. Como tales, tienen tres características que son importantes para el presente análisis: (1) en relación con los estándares y evaluaciones, el acuerdo sobre las metas es fácil de obtener; (2) las metas no siempre se alcanzan completamente y (3) siempre habrá metas educativas importantes, o aspectos de determinadas metas que, por distintas razones prácticas o conceptuales, no se medirán. Independientemente de lo cuidadosa que sea la especificación de las metas, no son, por naturaleza, estándares ni evaluaciones. El proceso de pasar de metas a actividades concretas de evaluación es característicamente complejo y abarca las numerosas cuestiones de los recursos y la ejecución. Así pues, en último análisis, las evaluaciones nunca representarán las metas deseadas con una cobertura completa o una fidelidad absoluta. No obstante, el proceso de establecimiento de metas es una empresa valiosa al igual que lo es un proceso bien concebido de evaluación.

Aun cuando resulta que muchas metas importantes no son directamente medibles,

la existencia de metas concretas sirve como función útil en la coordinación de las actividades y la planificación de la utilización de fuentes de datos indirectas y complementarias para seguir el proceso hacia el logro de las metas. Las metas explícitas también pueden ser útiles para resolver las diferencias acerca de políticas y prácticas concretas. Por ejemplo, gran parte del esfuerzo que ahora se está dedicando a mejorar la enseñanza de las matemáticas y las ciencias en los Estados Unidos emana de las comparaciones internacionales del desempeño de los alumnos en estas áreas con resultados desfavorables para los Estados Unidos. Estos datos permitieron a los educadores en los Estados Unidos proponer una meta de mejoramiento en su conjunto que fue aceptable para el público en general y, en el curso del tiempo, obtener recursos considerables para el largo proceso de fijar los estándares, crear las evaluaciones y proporcionar los recursos que se necesitarán para alcanzar esta meta.

Una vez que se ha llegado a un acuerdo sobre las metas, han de establecerse estándares concretos para alcanzarlas. Volviendo a mi ejemplo precedente, el progreso hacia una meta nacional consistente en mejorar la comprensión de las matemáticas y las ciencias por parte de los alumnos requiere establecer estándares curiculares en ambas disciplinas que especifiquen qué deberían aprender los alumnos y qué deberían enseñar los maestros y el nivel en el que debería demostrarse el rendimiento. En los Estados Unidos en nuestros días, este proceso incluye al menos dos perspectivas diferentes: las de cada uno de los estados que aplicarán los estándares y las de los profesionales de la educación, tales como las asociaciones nacionales de materias cuyos miembros en el campo de la docencia, formación de docentes e investigación tienen experiencia curricular apropiada. Los estándares para los maestros también han de formularse y coordinarse con los estándares para resultados de los alumnos, ya que es razonable suponer que los maestros no pueden enseñar con eficacia lo que ellos mismos no conocen. Los estándares sobre resultados para los maestros pueden tener repercusiones sobre el contenido de sus programas de formación de docentes. Los estándares sobre el contenido para los maestros se ocupan de la materia y el conocimiento pedagógico que los maestros deberán poseer y de su capacidad para utilizar este conocimiento en el aula. El insumo, o la oportunidad para aprender los estándares, asegura que los maestros tienen acceso a experiencias educativas continuas, materiales y otros recursos que ellos y sus alumnos necesitan para producir los resultados educativos requeridos.

En los estándares que actualmente se establecen en los Estados Unidos, han surgido varios temas importantes que probablemente sean aplicables a otros lugares. El proceso de establecimiento de estándares no sólo aclara lo que se ha de enseñar y aprender, sino que también eleva el nivel de expectativas acerca del desempeño. Los estándares establecidos en un proceso consensual, que incluya a los profesionales de la educación, tienden a ser refinados, exigentes y progresivos. Si bien la creación de estándares altos constituye un resultado altamente valorado del proceso de establecimiento de estándares en términos de creación de un clima de altas aspiraciones para aprendizaje de los alumnos, también crea cierto número de

problemas prácticos que hay que resolver. Los maestros, las organizaciones profesionales de maestros y los educadores de maestros estarán justificablemente preocupados sobre cómo alcanzarán estos altos estándares. Como cuestión práctica, estarán preocupados acerca de cómo obtener los recursos que pueden requerirse para ello.

Sin embargo, estas preocupaciones no han reducido su interés en formular y aplicar los estándares. Concretamente, muchos maestros no están preparados en la actualidad para enseñar al nivel que implican los estándares, por lo que se necesitará una extensa educación en el desempeño de las funciones para los maestros (y para otros educadores en sus programas de formación de docentes) a fin de cumplir los estándares. Además, si se va a esperar de los maestros en una amplia gama de escuelas y alumnos el mismo nivel alto de desempeño, hay que prestar atención a las disparidades en los recursos (Banks, 1997). Esto significa que los estándares han de preocuparse no sólo del producto final de la enseñanza de los alumnos (producto educativo), sino también de los recursos a disposición de los alumnos y maestros para ayudar en la enseñanza y el aprendizaje (estándares de insumos educativos u oportunidad).

Esta atención a los insumos puede parecer evidente en teoría (si se va a enjuiciar a las personas por lo que han aprendido, deberán tener la misma oportunidad para aprenderlo), pero la ejecución es muy controvertida. Entre las áreas de desacuerdo que han surgido a este respecto figuran quién debería tener la responsabilidad de asegurar la calidad de los insumos y cómo deberían proporcionarse los fondos. Además, han proliferado debates metodológicos en torno al tema de la oportunidad para aprender. Algunos aspectos de los estándares sobre insumos son relativamente fáciles de medir, tales como la disponibilidad de libros y la calidad de las instalaciones escolares. Sin embargo, otros aspectos muy importantes son mucho más difíciles de medir, tales como la calidad de los maestros. En los Estados Unidos, el establecimiento y la aplicación de estándares sobre insumos han sido muy controvertidos incluso en las etapas de planificación y promete seguir siéndolo, dada la amplia variación en la forma en la que se están conceptualizando estos estándares.

Debido a que las evaluaciones emanan directamente de los estándares, en principio deberían crearse sólo después de haber concluido la especificación de las metas y los estándares. En la práctica, no obstante, el proceso tiende a ser iterativo. Por ejemplo, a menudo la primera vez que las aspiraciones de estándares educativos se ponen en operación es cuando se traducen a tareas de evaluación o preguntas de prueba. Aun cuando esto es lamentable en algunos respectos, la situación puede tener aspectos positivos si existe cierta cantidad de flexibilidad en el sistema educativo. La experiencia con la redacción y el ensayo de evaluaciones puede proporcionar una idea de cómo deberían perfeccionarse los estándares, ya que las evaluaciones tratan de traducir los estándares a indicadores medibles del desempeño mientras que se aspira a preservar la intención original. Cuando los estándares se traducen a tareas y preguntas concretas para incluirlas en una prueba, entonces la

gente critica las tareas como demasiado difíciles, demasiado fáciles, irrelevantes, injustas o políticamente correctas o incorrectas. No obstante, esta crítica puede proporcionar una oportunidad para agudizar y concentrar el significado de metas, estándares, evaluaciones y sus interrelaciones.

Han de establecerse los vínculos entre todos los elementos que contribuyen al logro de metas educativas—alumnos, maestros y currículos—a fin de que las evaluaciones de los maestros tengan máxima eficacia. Es importante que estos vínculos se establezcan al nivel de metas y estándares, no sólo a nivel de evaluación. En esta etapa del establecimiento de estándares en los Estados Unidos, a pesar de la cuestión separada de las dificultades creadas por las numerosas diferencias en formato, idioma, nivel de especificidad e intención, se están identificando numerosos casos de discrepancias sustantivas más amplias que necesitarán resolverse: discrepancias entre los estándares de grupos profesionales, los de los estados y los nacionales; discrepancias entre lo que se está pidiendo a los alumnos que aprendan y a los maestros que preparen; y, lo que es aún más importante, discrepancias entre los requisitos de hoy y de mañana para maestros y alumnos.
Estándares

Existen numerosos tipos de estándares que son aplicables a la evaluación de los maestros y hay mucha confusión generada por la falta de atención a los diferentes fines a los que sirven. A continuación presentamos ejemplos de los distintos tipos de estándares.

Estándares para validez, calidad técnica y utilización

La calidad de las evaluaciones de los maestros puede enjuiciarse por los estándares técnicos que se aplican ampliamente a cualquier evaluación educativa o psicológica. La teoría moderna de la validez (por ejemplo, APA, 1985; Cole y Moss, 1989; Messick, 1989; Moss, 1992) subraya el amplio contexto en el que se formulan y utilizan las evaluaciones y proporciona una guía conceptual para considerar la validez de las evaluaciones de actividades complejas tales como la enseñanza. Este criterio de la calidad de la evaluación pone de relieve la importancia de demostrar los vínculos lógicos entre partes del sistema educativo y un examen sistemático de las consecuencias de realizar evaluaciones. Este criterio es ahora ampliamente aceptado por los especialistas en mediciones. La serie de estándares para evaluaciones a los que se hace referencia más ampliamente, *Standards for Educational and Psychological Testing* (Asociación Norteamericana de Investigaciones Educativas, Asociación Psicológica Norteamericana y Consejo Nacional de Medición de la Educación, 1985) refleja este punto de vista al igual que lo hace la revisión de estos estándares que ahora está en vías de realización (la fecha prevista para conclusión es 1998). No obstante, estos estándares son tan generales que dejan mucho a la discreción de quienes elaboran y utilizan las evaluaciones de los maestros. Entre las publicaciones adicionales sobre estándares y expectativas técnicas en esta esfera figura la publicación del Comité Conjunto sobre Estándares de Evaluación Educativa titulada *Personnel Evaluation Standards* (1988). Estos

estándares incluyen no sólo la atención a la calidad de las evaluaciones sino también a las características del contexto educativo y el empleo ético de las evaluaciones.

La Junta Nacional de Estándares de Enseñanza Profesional (NBPTS) también ha ofrecido una perspectiva sobre los estándares técnicos y otros específicos para la evaluación de la enseñanza que han tenido influencia en los recientes debates de las evaluaciones de los maestros. La NBPTS ha aplicado a sus propias actividades el estándar de que: "Las evaluaciones han de ser profesionalmente creíbles, públicamente aceptables, legalmente defendibles, administrativamente fiables y económicamente asequibles" (NBPTS, 1991, página 53).

Hay una creciente disponibilidad de los estándares propuestos relacionados específicamente con la calidad técnica de las evaluaciones del desempeño aunque ninguno ha alcanzado aún el nivel de aceptación general de los estándares de la AERA/APA/NCME o el Comité Conjunto. Algunos estándares propuestos relacionados con los distintos aspectos de la evaluación del desempeño de los maestros han sido ofrecidos por Claxton, Murrell y Porter (1987); Linn, Baker y Dunbar (1991); Moss (1992); Miller y Legg (1993); y Quellmalz (1991). Dwyer (1994) brinda un análisis adicional de esta literatura y su aplicación a las evaluaciones de los maestros.

Estándares curriculares

Los estándares curriculares se refieren a series de declaraciones de conocimiento y aptitudes que han de aprender los alumnos, presentados de ordinario en un marco que indica el alcance del contenido que se tratará y la secuencia en la que se ha de enseñar por lo general el material. Los estándares curriculares son pertinentes para las evaluaciones de los maestros en dos formas. Primero, series de currículos de los alumnos, tales como las elaboradas por el Consejo Nacional de Maestros de Matemáticas (1989) al definir lo que han de aprender los alumnos, tienen repercusiones directas para lo que han de conocer y ser capaces de enseñar los maestros. Por implicación, dichos estándares curriculares orientados al alumno también guardan relación con las actividades de desarrollo profesional de los maestros que se necesitan. Segundo, los estándares curriculares pueden formularse también para la formación de docentes. Dichos estándares se refieren directamente al contenido, al conocimiento y a las aptitudes pedagógicas específicos para el contenido que han de poseer los maestros. Para que dichos estándares sean eficaces deberían vincularlos con el conocimiento y las aptitudes que han desarrollarse en los alumnos.

Estándares para conocimiento y aptitudes de los maestros

Tal como se indicó arriba, los estándares curriculares pueden especificar directa o indirectamente el conocimiento y las aptitudes que se requerirán de los maestros. Los estándares para el conocimiento y las aptitudes de los maestros están vinculados conceptualmente con estos estándares curriculares, pero abordan la cuestión de

forma que (1) guarda relación directamente con los maestros y (2) especifica conocimiento y aptitudes que trascienden del conocimiento relacionado con el currículo de determinadas disciplinas. Los estándares para el conocimiento y las aptitudes de los maestros pueden abordar el conocimiento del contenido de la materia por parte de los maestros; el conocimiento por los maestros de cómo enseñar ese contenido (pedagogía específica para el contenido); el conocimiento por los maestros de principios pedagógicos generales (por ejemplo, desarrollo del idioma, recompensas y castigo); y la capacidad de los maestros para aplicar este conocimiento y estas aptitudes en un entorno de aula (desempeño pedagógico).

Según Bridges (1986), en la práctica, las escuelas evalúan a los maestros ya empleados de acuerdo a cinco criterios amplios: (1) conocimiento de la materia; (2) capacidad de impartir el conocimiento; (3) capacidad de mantener la disciplina en el aula; (4) capacidad de mantener un clima apropiado en el aula y (5) capacidad de establecer una relación con los padres y alumnos. No obstante, como regla general, el personal escolar no tiene una comprensión concreta del significado de estos criterios y, por tanto, con frecuencia los maestros no están seguros de lo que se espera de ellos o cómo mejorar su actuación. En el marco escolar, la determinación de la competencia de los maestros se realiza a menudo a través de las observaciones de los rectores u otro personal supervisor, aunque eso no constituya el mejor método para obtener información sobre todas las cinco áreas identificadas por Bridges. Es raro que el aprobar una prueba estandarizada sea uno de los criterios para seguir en el empleo pero así ha ocurrido en varias jurisdicciones en los Estados Unidos.

Las evaluaciones de los maestros para fines distintos de la evaluación anual de los ya empleados presentan diferentes opciones de evaluación. Se ha creado un ejemplo de un enfoque diferente para uso en certificar que los maestros cumplen estándares profesionales muy altos y son profesionales excelentes. La NBPTS ha diseñado evaluaciones complejas e integrales, para los maestros experimentados, que están vinculadas a niveles de edades concretas de los alumnos y a áreas de materias concretas. Las evaluaciones de la NBPTS son, característicamente, una combinación de métodos de evaluación (fuertemente orientados hacia la evaluación del desempeño) que son específicos para la enseñanza en un determinado contexto, tal como las aptitudes del lenguaje de los primeros años de la adolescencia. Las evaluaciones son externas en el sentido de que son enjuiciadas por evaluadores especialmente capacitados que son educadores experimentados y no forman parte del entorno de empleo del maestro. Las evaluaciones de la NBPTS pueden obtenerse ahora en ciertas áreas de materias básicas y edades que incluyen a un gran número de maestros; otras están aún en proceso de desarrollo. Además del desarrollo de evaluaciones, la NBPTS también ha realizado una gran cantidad de investigaciones en cierto número de áreas relacionadas con el uso de simulaciones para la evaluación de los maestros (por ejemplo, desempeño grabado en videocinta; el uso de portafolios de trabajo; ejercicios en centros de evaluación; y el impacto de las evaluaciones de la NBPTS sobre los maestros que participan en ellas).

Otro modelo, elaborado originalmente para la evaluación de los maestros que comienzan, pero que se ha ampliado ahora para incluir también a los maestros experimentados, es el de la Serie de Práctica elaborada por el Servicio de Pruebas Educativas (Dwyer, 1994; Dwyer y Ramsey, 1995; Dwyer y Villegas, 1993). La Serie de Práctica ofrece actividades de desarrollo profesional y evaluación que utilizan distintos métodos de evaluación para recopilar datos para fines de toma de decisiones personales e institucionales. Las áreas de evaluación son:

- Aptitudes básicas (lectura, escritura y matemáticas) requeridas de todos los posibles maestros para el éxito en la formación de docentes y buen desempeño posterior.
- Conocimiento de la materia y conocimiento pedagógico específico para el contenido. Estas evaluaciones utilizan métodos de recopilación de datos por escrito, basados en computadora, orales y de otra índole.
- Conocimiento de los principios pedagógicos básicos (utilizando métodos de recopilación de datos tales como los descritos arriba).
- Aplicación de conocimientos y aptitudes en el aula. Estas evaluaciones son predominantemente evaluaciones del desempeño y dependen de la observación en la propia aula del maestro, las entrevistas con el maestro y los documentos de recopilación de datos.

Estándares de nivel de rendimiento

Además de especificar el "qué" en los estándares de enseñanza y aprendizaje, también ha de abordarse la cuestión de "cuánto" Los estándares de conocimiento y aptitudes curriculares y del maestro abordan a menudo esta cuestión por implicación y con menos frecuencia la abordan directamente. Un ejemplo de abordar los estándares de nivel de conocimiento y aptitudes por implicación, incluido el cálculo como elemento de los estándares de matemáticas, implicaría de ordinario un nivel más elevado de resultados de desempeño que los estándares que especificaban sólo el contenido de matemáticas enseñado habitualmente antes del cálculo.

No obstante, la cuestión del nivel de estándares de desempeño se aborda de forma apropiada, independientemente de dichas consideraciones. Los estándares pueden establecerse en tres niveles (individualmente o en combinación). Por ejemplo, puede establecerse un estándar general para determinar, sobre una base de sí o no, si un maestro reúne las condiciones para enseñar o para que se le renueve un contrato de empleo. A un nivel más detallado, los estándares pueden establecerse para determinar si un maestro posee conocimiento adecuado de un aspecto en particular de la literatura o la capacidad de mantener la disciplina en su aula.

Dwyer (1997) proporciona un análisis general de los principios fundamentales utilizados para establecer dichos estándares. La naturaleza de las puntuaciones límite[2] no está en hallar la puntuación "correcta", sino en el enjuiciamiento de uno o más individuos habilitados o autorizados acerca de la cuestión "¿Cuánto es

suficiente?" La respuesta no está en la prueba o en el método utilizado para establecer la puntuación límite. Esta es una pregunta que sólo puede responderse con respecto a factores que trascienden de la propia prueba, con referencia al contexto en el que se utiliza la prueba y las percepciones de los jueces acerca de dicho contexto.

En las evaluaciones de los maestros, tales estándares pueden ser muy controvertidos. El mismo estándar puede ser considerado simultáneamente por personas diferentes como (1) tan alto que representa una barrera desmedida del empleo de personas cualificadas que se convertirían en maestros aceptables; o (2) tan bajo que presta un mal servicio a los alumnos al permitir a personas no cualificadas para enseñar. Dwyer advierte que esta situación no resulta de haber seleccionado la puntuación límite "incorrecta" o de deficiencias técnicas de la prueba en la que se utiliza la puntuación límite. Resulta de claros desacuerdos acerca de la importancia relativa de diferentes clases de error de clasificación y acerca de la importancia relativa de las políticas educativas tales como la disponibilidad de maestros, la elevación de los estándares educativos y las preocupaciones acerca de la equidad para maestros y alumnos.

Tres aspectos centrales y mutuamente relacionados acerca de tales estándares son pertinentes para la evaluación de los maestros:

1. Todos los métodos de establecimiento de puntuaciones límite dependen de juicios. Los juicios pueden efectuarse acerca de las personas o de las preguntas de la prueba (u otros aspectos de la propia prueba), pero el juicio con influencias independientes de la prueba es una característica intrínseca de cualquier puntuación límite.

2. El establecimiento de puntuaciones límite conducirá invariablemente a errores al determinar si una persona ha cumplido o no el estándar. Las puntuaciones límite imponen casi siempre diferenciaciones externas sobre una distribución continua. Muy pocas evaluaciones de cualquier clase que sean pueden distinguir de forma fiable entre personas con puntuaciones adyacentes, pero, al aplicar una puntuación límite se obliga en realidad a efectuar dicha distinción.

3. No existe una puntuación límite "verdadera" que pueda hallarse con la aplicación del método correcto o una muestra de jueces suficientemente grande. La determinación de una puntuación límite es similar a la estimación de un parámetro de población. El uso de muestras mayores de jueces, o jueces mejor capacitados, en estudios destinados a establecer puntuaciones límite mejorará ciertos aspectos técnicos concretos (por ejemplo, reducir el error de muestreo (Lawshe, 1975), pero esto no resultará en una puntuación límite que sea superior a las puntuaciones límite que pudieran haberse establecido con otra selección de jueces cualificados o con otra metodología igualmente justificable para la selección de la puntuación límite.

Estándares de oportunidad para aprender

En un contexto que se ocupa principalmente de estándares para los alumnos, Banks (1997) advierte la importancia de la oportunidad al especificar los estándares. Indica que, al establecer los estándares educativos nacionales *Goals 2000* en los Estados Unidos, entre las clases de estándares considerados figuraron estándares de contenido, de desempeño, de evaluación y de oportunidad para aprender. Los estándares de oportunidad para aprender fueron los más controvertidos de las cuatro clases. Estos últimos abordan condiciones en las escuelas y las comunidades que limitan la capacidad de los alumnos y maestros para alcanzar los otros tipos de estándares. Entre los ejemplos de dichas variables en cuanto a insumos figuran la calidad de las instalaciones escolares, la disponibilidad de materiales docentes y el nivel de experiencia pedagógica disponible en las escuelas. La Figura 1, que antecede, muestra el papel de los estándares de oportunidad en el proceso de mejoramiento de la educación. A pesar de la lógica para tener dichos estándares, las críticas que implican de los mecanismos actuales de apoyo de las escuelas los hacen con frecuencia poco apetecibles por razones políticas. Los detractores del empleo de dichos estándares argumentan que reducen el énfasis en los resultados de desempeño y la responsabilidad al concentrarse de forma inapropiada en los recursos y los insumos.

Cabe advertir también que, en el caso especial del *licenciamiento* de los maestros que comienzan a ejercer la profesión, el concepto de la oportunidad para aprender no es relevante. En los contextos de licenciamiento, la decisión que ha de hacerse es la de si el maestro posee el conocimiento y las aptitudes necesarios para ser considerado apto en situaciones de enseñanza concretas. Debido a que estas situaciones son aquellas en las que se hace hincapié en la protección del público contra el daño que puede producir un posible maestro incompetente, se pide al posible maestro que demuestre la posesión del conocimiento y las aptitudes estipulados. El método mediante el cual se obtuvieron esas aptitudes o las circunstancias para obtenerlas o no obtenerlas no es estrictamente pertinente en este contexto limitado. Si se desean antecedentes adicionales sobre esta cuestión, véase Shimberg (1985).

FINES DE LA EVALUACIÓN

El proceso integrado que acabo de describir de pasar de metas a evaluaciones puede coadyuvar al logro de cierto número de objetivos en la evaluación de los maestros. Lamentablemente, por lo general ocurre que la finalidad de la evaluación no es universalmente clara para todas las partes interesadas cuando se adopta inicialmente la decisión de crear evaluaciones de los maestros. En la práctica, a menudo hay diferentes puntos de vista acerca de la finalidad de las evaluaciones, aun después de haberse comenzado a utilizarlas. Un paso esencial en el diseño de la evaluación consiste en adquirir una clara comprensión de las razones para la evaluación y los usos que se harán de los datos de la misma, pero, equivocadamente, este paso se da a menudo por sentado. Así pues, es importante ser claros acerca de los fines

perseguidos con las evaluaciones de los maestros.

Fines de la evaluación

Diferentes puntos de vista sobre la finalidad de las metas de la mejora educativa entrañan a menudo diferentes puntos de vista de las evaluaciones de los maestros. Dichos puntos de vista sobre las evaluaciones de los maestros pueden ser bastante diferentes con respecto a la sustancia de las evaluaciones; los defensores y detractores de las evaluaciones; la razón fundamental para las evaluaciones; y los enfoques prácticos para las evaluaciones (Dwyer y Stufflebeam, 1996). Entre algunas de las metas principales de las evaluaciones de los maestros que con frecuencia se observan en la práctica y se citan en la literatura, figuran las siguientes:

- Mejoramiento de la enseñanza en el aula. Los educadores profesionales son importantes defensores de este punto de vista. Entraña un desarrollo educativo continuo a lo largo del cual una persona puede mejorar, una preferencia por evaluaciones formativas en vez de sumativas, y un fuerte vínculo con las actividades de desarrollo profesional. Si se desea un análisis adicional de esta materia, véanse Hunter (1988); Duke y Stiggins (1990); y Shulman (1986).

- Responsabilidad y desarrollo profesionales. Los maestros y sus asociaciones profesionales son los defensores principales de este punto de vista. Entraña una fuerte visión de la enseñanza como profesión con sus propios estándares, ética e incentivos intrínsecos para la persona consagrada. Aun cuando las responsabilidades son un elemento clave en este punto de vista, la responsabilidad es a la profesión y a sus estándares de práctica y ética, más que a entidades externas tales como empleadores o el estado. Véase un análisis adicional de esta materia en la Junta Nacional de Estándares Docentes Profesionales (1991).

- Control administrativo. Los administradores de escuelas son los defensores principales de este punto de vista. Entraña considerar la enseñanza como una situación de empleo que requiere supervisión y control del maestro por la unidad administrativa. En el ámbito de las escuelas públicas en los Estados Unidos, la base para este criterio se fundamenta en la protección del público contra práctica negligente. Para un análisis adicional de esta materia, véanse Andrews (1985) y Redfern (1963, 1980).

- Pago por mérito. En este punto de vista, que puede considerarse como subserie de la responsabilidad y el control administrativo que describimos arriba, se estima que los maestros necesitan el reconocimiento y la motivación que proporcionan los incrementos salariales. Los defensores principales de este punto de vista son el público en general y los funcionarios del gobierno que lo representan. Los defensores de este punto de vista desean a menudo utilizar el rendimiento de los alumnos como indicador del mérito para el que los aumentos

de sueldo sirven de recompensa. La aplicación de este criterio puede entrañar la creación de una escala profesional con una serie de pasos vinculados con el desempeño. Si se desea un análisis adicional de esta materia, véanse Webster, Mendro y Almaguer (1993).

Responsabilidad y mejora educativa

El seguimiento/responsabilidad, el mejoramiento y la instrucción y la evaluación programática son los tres fines más frecuentemente citados para la evaluación de los maestros en los Estados Unidos en nuestros días. Aunque están estrechamente relacionados entre sí conceptualmente, la evaluación para estos fines difiere bastante en diseño y en sus repercusiones en cuanto a la forma en que se utilizarán los datos obtenidos con estas evaluaciones. Por ejemplo, la evaluación para responsabilidad y evaluación programática puede utilizar una muestra de personas y preguntas o ejercicios de evaluación en vez de pedir a cada individuo que complete cada una de las partes de la evaluación. En contraste, las evaluaciones concebidas directamente para mejoramiento de la instrucción se administran de forma más apropiada a cada maestro o alumno y deberían estar estrechamente vinculadas con las áreas del currículo que son más importantes para estos alumnos y sus maestros.

La responsabilidad es, por tanto, no una meta en sí sino, más bien, un medio hacia el fin del mejoramiento de la educación. A menos que se planifiquen con cuidado las consecuencias del buen y el mal desempeño de los maestros y, a menos que el sistema de evaluación sea justo de hecho y en aspecto, la evaluación de la responsabilidad puede, en realidad, subvertir su mira pretendida de mejoramiento de la instrucción. El uso generalizado de las evaluaciones para responsabilidad de los maestros en los Estados Unidos ha creado gran cantidad de controversia. La investigación ha demostrado que, sin planes cuidadosos para recopilación y utilización de los datos, pueden obtenerse resultados que chocan directamente con sus fines originales. Por ejemplo, los maestros y administradores que no valoran el contenido de las evaluaciones pueden recurrir a medios ilícitos para obtener puntuaciones aceptables.

Surgen problemas tanto éticos como prácticos cuando el contenido de las evaluaciones no se vincula estrechamente con aspectos importantes de la enseñanza. En tales circunstancias, el tiempo empleado en la preparación de las evaluaciones se considerará con razón como tiempo perdido. Cuando estos vínculos están ausentes entre las evaluaciones y las metas y estándares educativos más amplios, hay también razón para preocuparse de que «se obtiene lo que se evalúa»—es decir, el tiempo y la atención dedicados a los temas que aparecen en las evaluaciones, a expensas de otros temas más importantes que no aparecen. Como cuestión a este respeto se cita a veces el formato de las evaluaciones. Una dependencia excesiva en una forma única de pruebas puede hacer que los maestros (o los alumnos) adquieran habilidad en la clase de reflexión que requiere esa forma de pruebas a expensas de otras clases.

Para establecer fuertes vínculos entre la responsabilidad de los maestros y el

mejoramiento educativo se requiere prestar atención a los mecanismos y recursos de que se dispone para resolver los problemas cuando los revelan los datos de las evaluaciones. La correspondencia entre las metas y los estándares educativos, por un lado, y el formato y el contenido de las evaluaciones de los maestros, por el otro, ayuda a los responsables de la política educativa y al público a comprender claramente dónde existen problemas, lo que a su vez puede sugerir sus soluciones.

Diferentes fines para las evaluaciones de los maestros implican diferentes estándares, métodos de evaluación y énfasis. Un factor importante en la evaluación de los maestros para fines relacionados con las decisiones administrativas, tales como la responsabilidad y el pago por mérito, es la existencia de escaseces reales o potenciales de maestros. Rara vez se hallarán útiles las evaluaciones que exacerban las escaseces de maestros en áreas vitales. Históricamente en los Estados Unidos, dichas escaseces han ocurrido en materias técnicas tales como las matemáticas, en las que la educación compite, de ordinario sin éxito, con la industria por personas cualificadas. Para análisis de éstas y otras cuestiones afines sobre la oferta y demanda de maestros, véanse Murnane y Schwinden (1989) y Sedlak y Schlossman (1986).

Consideraciones especiales para las evaluaciones de los maestros principiantes

Las evaluaciones de los maestros principiantes y experimentados difieren típicamente en términos de las metas que se pretende alcanzar con las evaluaciones y en la forma que adoptan estas últimas. Esto no quiere decir que no estén relacionadas ambas; un distintivo de una evaluación cuidadosamente planificada, de cualquier clase que sea, es el cuidado con el que se ha alineado con etapas previas o posteriores en la carrera del maestro. Una visión común del maestro debería regir la evaluación y su contenido en todas las etapas.

Las evaluaciones de los maestros que comienzan su carrera tienen de ordinario el fin primordial de proteger los intereses de los alumnos y del público de los efectos nocivos de una docencia subestándar. Dichas evaluaciones están orientadas característicamente a asegurar que el posible maestro posee el conocimiento y las aptitudes que le cualificarían para empleo como maestro pero, de ordinario, no garantizarían a la persona un cargo como maestro (Madaus y Mehrens, 1990; Mehrens, 1987; Rebell, 1990). El conocimiento y las aptitudes abarcados se consideran indispensables pero no necesariamente suficientes para fines de determinar la idoneidad de un cargo docente en particular.

En los Estados Unidos, las licencias de los maestros se otorgan típicamente bajo la autoridad de un gobierno estatal o una entidad que ha sido habilitada para desempeñar esta función.

Las evaluaciones de la Práctica para el licenciamiento de los maestros que comienzan su carrera son un ejemplo de las pruebas impartidas para tal finalidad.

En contraste con las evaluaciones de maestros experimentados, las evaluaciones de los maestros principiantes abarcan aptitudes habilitadoras de lectura, escritura y matemáticas. El conocimiento de la materia y la pedagogía específica para el contenido en las evaluaciones de los maestros principiantes concuerda con las expectativas de los maestros experimentados. La suficiencia en la aplicación del conocimiento y las aptitudes en el aula es de naturaleza análoga a la que se espera de los maestros experimentados, pero a un nivel inferior.

Un aspecto crítico de la evaluación de los maestros que comienzan su carrera es el de que las evaluaciones se utilizan para adoptar una decisión dicótoma. Aunque sirven también para otros fines, las evaluaciones de los maestros principiantes resultan característicamente en una decisión afirmativa o negativa acerca de la capacidad del posible maestro para avanzar en la profesión docente. Una implicación de medición de esta característica de las evaluaciones de los maestros principiantes es la de que la evaluación ha de estar concebida para facilitar esta decisión dicótoma, en vez de describir graduaciones de desempeño igualmente bien a lo largo de todo el continuo del conocimiento y aptitudes docentes. En la práctica, esto significa que los diseñadores de la evaluación han de comprender las cualidades que diferencian al candidato aceptable del candidato no aceptable y proporcionar tareas de evaluación que permitan con exactitud hacer diferenciaciones en el desempeño a este nivel en particular, en vez de a través de una gama más amplia de desempeño.

Crítica de la evaluación de los maestros

La evaluación de los maestros ha sido un área de las pruebas educativas especialmente controvertida casi desde sus comienzos a principios de la década de 1930. Las críticas de las evaluaciones de los maestros han diferido durante los años y con fines y métodos diferentes de evaluación.

En general, las críticas de la evaluación de los maestros principiantes para entrada en la profesión han subrayado a menudo la imparcialidad de los instrumentos y procedimientos, en especial con respecto al desempeño de los posibles maestros que son miembros de grupos minoritarios (por ejemplo, Darling-Hammond, 1986; García, 1985; Haney, Madaus y Kreitzer, 1987; Hood y Parker, 1991), alegando que las prácticas de evaluación de los maestros se llevan a cabo a menudo de forma poco uniforme que infringen los derechos de quienes son evaluados. Las críticas basadas en el daño hecho a los alumnos por la falta de conocimientos docentes del maestro son bastante raras en esta esfera.

Las críticas de la evaluación de los maestros en el ejercicio de su profesión para fines de responsabilidad, pago por méritos, etc., también han suscitado preocupaciones acerca del impacto real o potencial de las prácticas de evaluación de los maestros sobre los derechos de los posibles maestros a licencias docentes o acceso a empleo; sobre los currículos de las instituciones de formación de docentes; y sobre el ejercicio de la propia profesión docente. La mayor parte de las críticas que aparecen

en la literatura profesional guarda relación con el uso de pruebas de opciones múltiples para el conocimiento de la materia. Desde fines de la década de 1980, con el mayor uso de las evaluaciones de desempeño tanto para maestros como para alumnos, las críticas en la literatura profesional han tendido cada vez más a concentrarse en los aspectos prácticos de las evaluaciones del desempeño. Dwyer y Stufflebeam (1996) han identificado ocho cuestiones principales en las evaluaciones de los maestros que guardan relación principalmente con el uso de las evaluaciones para la selección y el empleo de los maestros:

- El uso de sistemas de evaluación no validados.
- El uso insuficiente de estándares profesionales para la planificación y el mejoramiento de los sistemas educativos.
- Opciones ineficaces de criterios claros, válidos y aplicables para determinar el desempeño de los maestros.
- La necesidad de técnicas y materiales para realizar los pasos básicos de la evaluación de los maestros.
- La falta de capacitación en evaluación para los evaluadores.
- La falta de guía sobre la evaluación de maestros de poblaciones especiales, tales como alumnos con impedimentos o los que necesitan educación bilingüe.
- La no consideración del contexto del aula y la escuela en la realización de las evaluaciones.
- La ausencia de una base teórica para las evaluaciones.

A pesar del apoyo profesional generalizado para la validación de las evaluaciones de los maestros (por ejemplo, Linn y colaboradores, 1989; Madaus, 1990), la mayoría de las evaluaciones de los maestros utilizadas por las escuelas no han sido rigurosamente validadas (Burry, Chisholm y Shaw, 1990; Seriven, 1987; Streifer e Iwanicki, 1987).

En una crítica relacionada con las dos últimas cuestiones de esta lista, algunos detractores, reconociendo la complejidad de la enseñanza, están específicamente preocupados con la factibilidad de determinar el conocimiento pedagógico junto con la comprensión de las materias. Las preguntas de opciones múltiples que tratan de evaluar la aplicación del conocimiento pedagógico fracasan en la cuestión del contexto: lo que constituye buena docencia varía con la materia que se está enseñando y con los antecedentes y las características individuales de los alumnos a los que se enseña. En ausencia de información muy detallada acerca del contexto del aula, que es imposible en términos prácticos en el formato de opciones múltiples, no se puede llegar a una conclusión con certeza acerca de lo que constituye instrucción apropiada. Las evaluaciones de las actitudes docentes, en contraposición al conocimiento necesario de las materias, necesita por tanto depender de formas más complejas de recopilación de datos.

Las evaluaciones del desempeño de los maestros principiantes creadas por los estados, aunque son nuevas en relación con las pruebas de opciones múltiples, no

han escapado a su porción de críticas. Dichos sistemas se han basado en observaciones de baja inferencia del comportamiento de los maestros; es decir, la tarea de observación se limita a comportamientos claramente observables que requieren un bajo nivel de inferencia por parte de los evaluadores y, por tanto, puede preverse que presenten un alto nivel de acuerdo entre evaluadores. Aunque este alto nivel de acuerdo es una característica deseable de las calificaciones, no es tan deseable que se tenga que obtener en aras de la validez de las calificaciones. Las calificaciones de baja inferencia, por su propia naturaleza, no pueden hacer frente a las facetas importantes pero inobservables de la enseñanza, tales como la toma de decisiones del maestro. Muchos detractores de las evaluaciones de los maestros basadas en el desempeño han indicado la deficiencia de los sistemas de baja inferencia—tanto los basados en las simulaciones como en las interacciones directas en el aula—que no toman en cuenta el contexto del aula como parte de la evaluación.

Otras críticas de la evaluación de los maestros guardan relación con el impacto de la evaluación de los maestros sobre los currículos de formación de docentes (Milner, 1991) y sobre la propia profesión pedagógica (Darling-Hammond, 1986). La preponderancia de esta crítica concierne al empleo de pruebas estandarizadas de opciones múltiples del conocimiento de la materia y pedagógico, más que a las evaluaciones del desempeño relacionadas con la capacidad de aplicar este conocimiento en el aula. Una cuestión importante para comprender las críticas de las evaluaciones de los maestros, y para formular políticas que eviten críticas legítimas, es hacer una distinción uniforme entre las dos clases de pruebas. Lamentablemente, las evaluaciones de las aptitudes básicas de habilitación (lectura, escritura y matemáticas) y el conocimiento de la materia, la pedagogía específica para el contenido y las aptitudes pedagógicas generales se tratan con demasiada frecuencia como si fuesen intercambiables con las evaluaciones del desempeño en el aula. Las inferencias hechas sobre la base del conocimiento de un maestro atañen a una condición necesaria *pero no suficiente* para una enseñanza eficaz en el aula. Si no se hace esta distinción simple pero vital se oscurecerán muchos debates acerca de las políticas y prácticas de evaluación de maestros.

Muchas de las deficiencias documentadas de las evaluaciones de los maestros pueden atribuirse a la falta de recursos, tales como procedimientos de capacitación adecuados para los evaluadores, instrumentos de evaluación e información procedente de estudios de validación. Sin embargo, otras críticas son el resultado del estado actual de la teoría y práctica de la evaluación de maestros. La ausencia de bases teóricas es un problema importante para muchos sistemas al evaluar tanto a los maestros principiantes como a los experimentados (Scriven, 1988a, 1988b, Wise, Darling-Hammond, McLaughlin y Bernstein, 1984). Lamentablemente, la literatura sobre investigación de la docencia ofrece poca orientación sobre cómo utilizar su teoría para formular y aplicar mejores evaluaciones de los maestros (Dwyer, 1994).

DEFINICIÓN DE LA ENSEÑANZA

Un reto importante en la creación de una serie bien formulada de evaluaciones de los maestros es la tarea de determinar qué constituye una buena enseñanza y qué estándares deberían utilizarse para determinar si se ha medido ésta de forma eficaz. La idoneidad de medir qué debería enseñarse no puede determinarse significativamente sin formular estándares explícitos, en un marco de referencia integral que abarque cuestiones tradicionales de preocupación para la educación, psicología y medición. Esta evaluación también ha de estar enmarcada en un contexto en relación con un determinado punto de vista de la enseñanza y el aprendizaje. Todos estos factores constituyen una parte apropiada de un concepto de la enseñanza que se utiliza para orientar decisiones posteriores acerca de los estándares y las evaluaciones. Las conclusiones acerca de la calidad de las evaluaciones, a su vez, han de guardar relación con el concepto de la enseñanza y los fines de las evaluaciones. Algunos acontecimientos recientes en el campo de las mediciones apoyan decididamente este punto de vista.

La teoría moderna de la validez, en su énfasis en un contexto amplio para el establecimiento de la validez de las evaluaciones, proporciona orientación conceptual al considerar la validez de las evaluaciones en desempeños complejos tales como la docencia y en los aspectos técnicos de la determinación del contenido. Dicho trabajo sobre la teoría de la validación pone de relieve las repercusiones para las mediciones de las interconexiones que existen en el seno de todo el sistema, del cual la evaluación es una parte, y el valor inherente en la medición directa del desempeño cuando esto es factible. Con el énfasis en el contenido y las consecuencias, las evaluaciones se examinan para determinar en términos amplios su beneficio para el sistema educativo del que forman parte. Por ejemplo, en la evaluación de los maestros, el contexto amplio de la evaluación de un maestro que comienza a enseñar incluiría factores tales como el distrito escolar, el estado y las políticas federales que afectan al ejercicio por el maestro de su profesión; el marco de la escuela y aula en el que ocurre la enseñanza; los recursos materiales a disposición del maestro; la naturaleza del currículo y la lección que se ha de enseñar; las características y el conocimiento previo de los alumnos; y la calidad de la preparación del maestro para la enseñanza.

Las consecuencias del empleo de las evaluaciones de los maestros se concentrarían análogamente en el impacto del empleo y resultados de la evaluación sobre la política educativa (¿Se crean nuevas políticas, basadas en los resultados de la evaluación? En caso afirmativo, ¿puede demostrarse que estas nuevas políticas son beneficiosas para los alumnos?); los materiales e instalaciones a disposición de los maestros y alumnos (¿Se identifican las necesidades materiales a través de la evaluación y, luego, se satisfacen dichas necesidades? ¿Desvía el proceso de evaluación la atención de las necesidades materiales de forma que se agravan dichas necesidades?); la calidad de la preparación de los maestros (¿Proporciona la evaluación información suficiente para los programas de formación de docentes a

fin de ayudar a los maestros a mejorar su programa?); la calidad del currículo (¿Conduce el proceso de evaluación y los resultados a la identificación de áreas del currículo que funcionan bien y otras que no funcionan bien?); y el aprendizaje de los alumnos (¿Conduce la evaluación de los maestros a un aprendizaje mejorado o interfiere con el desempeño en el aula de forma que los estudiantes aprenden menos?).

Si bien ahora hay una base teórica ampliamente aceptada para un criterio amplio y generalizado de la validez, incluyendo su énfasis en las consecuencias de la evaluación, y en realidad está codificado en la revisión más reciente de *Standards for Educational and Psychological Testing* (Asociación Norteamericana de Investigación de la Educación, Asociación Psicológica Norteamericana y Consejo Nacional sobre Medición en la Educación, 1985), hay aún un vacío considerable entre la literatura sobre validez y la serie diversa e interesante de cuestiones que se afrontan en la formulación de evaluaciones de los maestros, en particular cuando tienen lugar en ambientes en los que es mucho lo que está en juego y cuando se utilizan medios múltiples de recopilación de datos. Por ejemplo, los investigadores y elaboradores en esta esfera carecen de guía definitiva de la literatura sobre cuestiones tan importantes como las variables humanas y la materia en el aula en las evaluaciones de los maestros (cómo hacer frente a diferencias contextuales); y sobre las repercusiones para la validez de crear evaluaciones que toman como punto de partida el criterio de que el aprendizaje es un proceso activo consistente en construir significado a partir de experiencias previas (evaluaciones con una base constructivista). Estas cuestiones son vitales para determinar los criterios mediante los cuales se evaluará el desempeño docente pero se requiere llenar un vacío para pasar de la base teórica sólida en la literatura de investigación a la práctica real de investigación y desarrollo.

El cúmulo de publicaciones sobre validez, estándares, expectativas y aspiraciones—si bien pone de relieve importantes cuestiones conceptuales y prácticas—crea una pesada carga de interpretación y extrapolación para los responsables de formular las evaluaciones de los maestros, en particular quienes incluyen la atención al desempeño real en el aula. Es evidente que existe un acuerdo general en estas publicaciones en el sentido de que los responsables de elaborar las evaluaciones y los evaluadores deberían tomar en cuenta las consecuencias amplias de las evaluaciones; incorporar elementos del contexto en el proceso de evaluación; concentrarse explícita y exactamente en el conocimiento y aptitudes acerca de los cuales se desea sacar inferencias; e incluir en las evaluaciones toda la gama de contenido acerca de qué inferencias deben sacarse. A pesar de este alto nivel de acuerdo en principio, necesariamente se dejan a cada uno de los creadores de las evaluaciones y a quienes evalúan sus esfuerzos una gran cantidad de discreción y responsabilidad. La base para sus decisiones, si se considera cuidadosamente, puede ser un recurso importante para alinear las evaluaciones de los maestros con otros elementos del sistema educativo a fin de mejorar los resultados de la educación y alcanzar otras metas importantes.

Considerados como proceso, los elementos principales en la definición de la enseñanza para fines de evaluación son, entre otros, los siguientes:

- Determinar un concepto rector de la enseñanza.
- Formular un plan metodológico integral para definir la enseñanza.
- Vincular la definición de la enseñanza con las prácticas de evaluación.

Concepto rector de la enseñanza

Es de vital importancia para la formulación de evaluaciones eficaces de los maestros formular un concepto que guíe la enseñanza y el aprendizaje el cual reconozca explícitamente la conexión entre enseñanza y aprendizaje. Es imposible debatir qué es fundamental acerca de la primera sin considerar el segundo.

Un concepto eficaz que oriente la enseñanza debería poseer las siguientes características:

- Ser explícito y claro acerca de dónde proceden los criterios, con qué valores están de acuerdo y qué posiciones de valor rechazan.
- No confundir temas que son inherentemente cuestiones de estilo docente con los que son cuestiones importantes de sustancia (Scriven, 1988a, 1988b).
- Conducir a inferencias acerca del contenido de las evaluaciones y los métodos utilizados para recopilar los datos.

Por estas razones, el concepto que guíe la enseñanza deberá exponerse en una etapa inicial del proceso de desarrollo, antes de adoptar las decisiones finales de diseño.

Nuevamente, tomaré la Serie de Práctica como ejemplo para la formulación de un tal concepto rector. Este concepto ha sido formalizado por Dwyer y Villegas (1993). En el desarrollo de la Práctica, se hizo hincapié en la toma de decisiones por el maestro y en la importancia del contexto estudiantil, escolar y curricular al evaluar esa toma de decisiones. Este punto de vista implica decididamente el valor de la recopilación de datos en el marco real del aula en contraposición a simulaciones. Una segunda repercusión es la de que las evaluaciones deberían incluir oportunidades para que el evaluador y el evaluado interactúen en relación con el evento docente que se esté considerando a fin de que el evaluador comprenda plenamente las decisiones en las que se fundamentaron las acciones observables del maestro. Una tercera implicación es la de que, al no haber "una respuesta correcta única" a la pregunta de qué constituye buena enseñanza (debido a que la enseñanza se considera como inherentemente sensible al contexto), la puntuación de las evaluaciones ha de permitir formas múltiples de respuestas «aceptables», en tanto explican claramente lo que constituye práctica profesional inaceptable.

Una cuarta implicación relacionada con el concepto de la enseñanza como actividad cognoscitiva compleja es la de que el proceso de evaluación, dependiendo,

como lo hace, de la consideración de una serie compleja de datos, requerirá un enjuiciamiento profesional sustancial para su ejecución. Por lo tanto, los evaluadores habrán de ser profesionales experimentados que han recibido capacitación para alcanzar una comprensión común de los criterios de evaluación y otras consideraciones para su aplicación.

Este concepto de la enseñanza y el aprendizaje, y los criterios que emanan de él, tienen fuertes vínculos con las teorías y prácticas psicológica, educativa y de medición. Especifica un desempeño objetivo cognoscitiva y comportamentalmente complejo y proporciona un marco para examinar el impacto de las evaluaciones sobre el sistema educativo del que forma parte (alumnos; maestros que se están evaluando; la profesión docente; la formación de maestros y la capacitación del personal).

En el caso de la Práctica, también se identificaron explícitamente los valores siguientes como parte de la formulación del concepto rector y se utilizaron como guía para la formulación de la evaluación propiamente dicha:

- Compromiso al tratamiento equitativo de los maestros.
- Estándares y técnicas de evaluación que aborden las acciones y toma de decisiones de los maestros.
- Especificación de ciertas prácticas docentes como inaceptables en cualquier contexto, en tanto se permiten muchas modalidades diferentes de prácticas aceptables, permitiendo la creación de estándares de conocimiento y prácticas docentes específicas y significativas.
- Creación de una experiencia de aprendizaje positiva tanto para el maestro que está siendo evaluado como para el evaluador.

Especificación de una metodología para definición de la enseñanza

La Figura 2 presenta una metodología prototipo para definir el campo de la enseñanza para el desempeño en el aula por parte del maestro que comienza su carrera.

Esta metodología se utilizó en la creación de las Evaluaciones de la Serie de Práctica. Adviértase que éste es un proceso iterativo que integra datos sacados de la teoría mediante estudios de investigación formales y análisis de la literatura; y práctica mediante encuestas, participación y trabajo de campo iterativo de los maestros.

En la siguiente sección se describen cuestiones importantes que entraron en juego en la aplicación de esta metodología y en la formulación de evaluaciones de los maestros tomando como base este análisis.

Figura 2. Práctica III: Evaluaciones del desempeño en el aula. Esquema de formulación de criterios

CUESTIONES RELACIONADAS CON LA FORMULACIÓN DE EVALUACIONES DE LOS MAESTROS

Selección de los métodos de evaluación

La mejor estrategia para seleccionar los métodos de evaluación de los maestros puede resumirse muy simplemente: identificar primero qué es importante conocer acerca del maestro y, luego, determinar el procedimiento óptimo para obtener la información necesaria. Las posibilidades incluyen pruebas de opciones múltiples, otras formas de evaluaciones escritas y muchos tipos de determinación del desempeño. En la mayoría de los casos, una cobertura adecuada de los conocimientos y aptitudes necesarios requerirá una combinación de técnicas de evaluación. Los criterios para seleccionar el método óptimo para evaluar áreas específicas de contenido y aptitudes deberían ser integrales e incluir lo siguiente:

- Concordancia con el concepto rector de la enseñanza.
- Relación lógica con el área objetiva de conocimiento o aptitud.
- Impacto educativo positivo.
- Factibilidad operativa y económica.
- Imparcialidad demostrable.

Tal como indicamos antes, las evaluaciones básicas y de alta calidad deberían resultar en un efecto educativo positivo tal como se expresa en el criterio de Messick (1989) sobre validez de construcción. De acuerdo con este criterio, una prueba válida conduce a buena práctica educativa y aprendizaje fructífero y una prueba inválida conduce a una adquisición subóptima de aptitudes.

La imparcialidad para con los maestros que están siendo evaluados es un criterio clave para enjuiciar cualquier posible metodología de evaluación. La imparcialidad puede considerarse sobre la base de características tales como el género, la raza o la etnicidad; o sobre la base de otros factores tales como los antecedentes educativos, experiencias pertinentes de la vida y familiarización con la materia y los alumnos a los que se va a enseñar.

Otras características relacionadas con la imparcialidad para el individuo son la fiabilidad de las evaluaciones, que han de enjuiciarse por medidas apropiadas para cada método empleado en la evaluación. Diferentes formas de evaluación requieren diferentes métodos para establecer la fiabilidad así como diferentes estándares para designar niveles apropiados de fiabilidad. Concretamente, los métodos y niveles empleados para enjuiciar pruebas de opciones múltiples no son necesariamente apropiados para aplicación a otras formas de evaluación.

Una consideración importante al seleccionar un método de evaluación es la de que el método ha de ser operativo y económicamente factible para uso en gran escala.

La economía y eficiencia a corto plazo de las pruebas de opciones múltiples son evidentes. Sin embargo, también es aconsejable, al estimar nuevos métodos de evaluación, comparar el impacto total de las distintas formas de evaluación. Dentro de límites cuidadosamente definidos, el valor agregado de los aumentos en la validez, el carácter directo de la medición, la motivación de los candidatos y otras medidas de calidad de la evaluación pueden contrarrestar los mayores costos y complejidad operativa de otras formas de evaluación. Estándares de calidad de pruebas centrales tales como la fiabilidad, la validez y la imparcialidad, son aún pertinentes para las distintas formas de evaluación que se seleccionen.

La economía de las evaluaciones del desempeño y las pruebas de respuestas construidas en general presentan retos formidables. En comparación con las pruebas de opciones múltiples, estos métodos de evaluación son inherentemente métodos de alta densidad de mano de obra y, por tanto, son más costosos de poner en práctica. Por esta razón, lo que se mide con los métodos de respuestas construidas y evaluaciones del desempeño ha de ser más significativo para justificar los gastos adicionales requeridos. Las pruebas de respuestas construidas, tomadas en su conjunto, sirven a menudo de experiencia educativa para quienes efectúan la evaluación y son los objetos de ella. Este criterio de la evaluación como actividad en el desempeño de las funciones es a menudo parte de la razón fundamental para los gastos adicionales de las evaluaciones del desempeño.

En el caso de las evaluaciones de la Práctica, se seleccionaron distintos métodos de evaluación. La Práctica I, Evaluaciones de las Aptitudes Básicas de Lectura, Escritura y Matemáticas, se utiliza para las personas que entran en la profesión docente evaluadas mediante una combinación de preguntas basadas en computadoras, de opciones múltiples, abiertas y ensayos. La Práctica II, Evaluaciones del Conocimiento de la Materia, difiere en cuanto a la metodología según la disciplina. Casi todas las pruebas de la Práctica II contienen algunas preguntas de opciones múltiples y abiertas, pero ciertos aspectos difieren mucho. Una evaluación de los maestros de idiomas, por ejemplo, utiliza modalidades de respuesta audiovisual para evaluar los aspectos receptivos y productivos de la competencia en comunicación. Una evaluación de los maestros de educación física utiliza estímulos de vídeo para evaluar aspectos importantes de los movimientos corporales. La Práctica III, Evaluaciones del Desempeño en el Aula, utiliza una combinación de entrevistas, observaciones y documentación escrita para determinar la eficacia del maestro en un entorno de clase real.

¿Adelanto o atraso?

Los elaboradores de sistemas de evaluación de maestros también hacen frente a lo que se denomina a menudo el dilema de "adelanto/atraso". En imparcialidad para con los maestros que están siendo evaluados, los criterios empleados para enjuiciarlos han de reflejar la práctica profesional actualmente aceptable. Esto asegura también que las evaluaciones concuerdan lógicamente con los fines de éstas. Sin embargo, un valor competidor es el que, en vista del largo tiempo requerido

para formular evaluaciones de alta calidad y la probabilidad de que continúen siendo utilizadas por cierto número de años, también es importante no crear evaluaciones que, en realidad, queden anticuadas antes de que se concluyan o que fomenten una continuación de prácticas docentes que incluso ahora son marginalmente aceptables para la profesión.

Es central para esta cuestión el hecho de que la Práctica profesional actualmente aceptable no es un concepto estático y que casi diariamente se crea nuevo conocimiento acerca de la docencia. Al evaluar, si puede considerarse que un determinado aspecto de la enseñanza es apoyado por la investigación, es por tanto necesario hacer cierto número de enjuiciamientos complejos acerca del estado de la investigación y tomar en cuenta el consenso profesional acerca de tendencias futuras en esa esfera. También es necesario considerar cuán definitiva es a la sazón un área de investigación. Por ejemplo, el área del comportamiento de los maestros y sus vínculos con el aprendizaje de los alumnos ha sido objeto de extensa investigación durante muchos años. En determinadas subáreas, los campos están bien trazados. Son numerosos los estudios bien concebidos y, en algunos casos, se ha llegado a conclusiones definitivas.

En contraste, el área cognoscitiva de los maestros y sus vínculos con el aprendizaje del alumno es relativamente nueva y se encuentra en estado de evolución. Aunque, por lo general, no se disputa la importancia de este campo de la investigación para la práctica docente, la naturaleza cognoscitiva del maestro sí es aún algo incierta. Cierto número de principios importantes, aunque lógicamente bien expresados y convincentemente bien demostrados en los mejores estudios de investigación, no han sido reproducidos aún ampliamente. Para la elaboración de evaluaciones de los maestros es necesario efectuar enjuiciamientos basados en dichas observaciones acerca de qué incluir para los fines de evaluaciones de los maestros. Un factor que complica el problema es el de que la investigación sobre el comportamiento de los maestros y sobre el aspecto cognoscitivo de ellos tiende a utilizar diferentes metodologías, creando así otra dificultad en la evaluación de la investigación más nueva conforme a estándares tradicionales.

En el caso de la elaboración de las evaluaciones de la Práctica, la mayor parte del contenido se sacó de la práctica actual, pero ciertos aspectos se enjuiciaron como parte de tendencias actuales. Por ejemplo, al determinar el desempeño en el aula para la Práctica III, se hizo mayor hincapié en la importancia de que los maestros trabajen entre sí y con las familias de los estudiantes que el que las encuestas indican que es la práctica actual. La inclusión de esta faceta de la enseñanza se basó en el consenso profesional de que dichas interacciones son beneficiosas para el aprendizaje de los alumnos y en la investigación que apoya este hecho.

Énfasis prácticos y teóricos

Otro dilema que afrontan muchos elaboradores de cualquier sistema de evaluación de maestros es el mérito relativo del conocimiento teórico y práctico de la

enseñanza. Sternberg y Wagner (1993) han hecho una distinción útil entre los problemas académicos y los prácticos que puede ser interesante con respecto a la evaluación de los maestros. De acuerdo con Sternberg y Wagner, los problemas académicos tienden a (a) ser formulados por otras personas, (b) estar bien definidos, (c) ser completos con respecto a la información requerida para resolverlos, (d) tener sólo una respuesta correcta única, (e) tener sólo un método único para obtener la respuesta correcta, (f) no estar incorporados en la experiencia ordinaria y (g) tener un valor intrínseco escaso o nulo. En contraste, los problemas prácticos tienden a (a) requerir reconocimiento y formulación del problema, (b) estar mal definidos, (c) requerir la búsqueda de información, (d) tener soluciones aceptables múltiples, (e) permitir múltiples caminos que lleguen a la solución, (f) estar incorporados en la experiencia cotidiana y requerir dicha experiencia y (g) exigir motivación y participación personal (página 2). Es claro que, de acuerdo con Sternberg y Wagner, las evaluaciones integrales de maestros abordan los problemas prácticos que han de resolver estos últimos, más que los problemas académicos.

Además, tal como se indicó antes, es importante, por razones tanto prácticas como teóricas, que los criterios sobre los cuales se fundamentan las evaluaciones de los maestros y su marco de organización, guarden relación con la propia comprensión que tienen los maestros de su trabajo. Sin embargo, al propio tiempo, los criterios también deberán basarse en la teoría educativa y psicológica para establecer un marco lógico. También mostraron una generalización mayor a través de los contextos docentes y una mayor probabilidad de pasar la prueba del tiempo en el uso real en el aula. Nuevamente, la resolución de este dilema no es un proceso mecánico. El buscar los aportes de los maestros que ejercen su profesión y los teóricos de la educación para repasar y revisar los materiales de evaluación hasta que se perciba ampliamente que son aceptables desde ambos puntos de vista ha sido una estrategia eficaz en la formulación de las evaluaciones de la Práctica.

¿Desde qué perspectiva debería considerarse la base de conocimientos?

Tal como se indicó antes, las fuentes principales de datos para la base de conocimientos acerca de la enseñanza son, entre otras, los maestros en el ejercicio de su profesión y la perspectiva de investigación/teórica. Las fuentes adicionales pueden consistir en disposiciones que sustituyen a las escuelas rectoras y a los maestros. Cada una de estas perspectivas no proporciona simplemente un punto de vista diferente del mismo fenómeno; cada una formula preguntas diferentes, emplea métodos diferentes para llegar a conclusiones y tiene una serie diferente, aunque a menudo relacionada, de preocupaciones acerca del significado y la utilización del conocimiento sobre la enseñanza. Estos puntos de vista representan fundamentalmente paradigmas diferentes, en el sentido de que las hipótesis, las metodologías y los valores básicos difieren.

Por tanto, no es un proceso algorítmico o mecánico para llegar a criterios que incorporan datos procedentes de estas tres fuentes. Como ejemplo de una solución a este problema, los elaboradores de la Serie de Práctica realizaron un procedimiento

iterativo consistente en crear proyectos de criterios y presentarlos luego para análisis a representantes de esos tres puntos de vista principales. Se preguntó a los analizadores y miembros del grupo, en esencia, si los criterios representaban la base de conocimientos para la enseñanza como ellos la veían. Los criterios se sometieron a cierto número de importantes revisiones como resultado de este proceso. Con cada una de estas importantes revisiones, se emprendieron ciclos de trabajo de campo cada vez más largos para proporcionar datos adicionales. Los criterios de la Práctica III se consideraron definitivos cuando cada una de las partes interesadas a quienes se consultó, quedó satisfecha de que se habían abordado sus preocupaciones principales.

Alcance de los criterios de evaluación

Necesitan abordarse dos temas interrelacionados que pudieran caracterizarse como cuestiones relacionadas con el ámbito de los criterios en el proceso de traducir el conocimiento del campo de la enseñanza a criterios específicos de evaluación del maestro: determinar el «tamaño» o nivel apropiado de generalidad para los criterios; y determinar la gama de contextos de enseñanza a los que se aplican los criterios.

La enorme variabilidad entre los contextos del aula (Shulman, 1988a, 1988b; Stodolsky, 1988) presenta retos importantes para toda actividad de evaluación de los maestros. En el caso de la Práctica III, Evaluaciones de Desempeño en el Aula, los criterios de enseñanza se diseñaron como *aspectos* de ésta; es decir, como principios que deberían aplicarse en una amplia gama de contextos docentes (incluida la variabilidad en la materia y el nivel de grados enseñado, el estilo docente y las características individuales y antecedentes de los alumnos) en vez de como «reglas» específicas que deberían seguirse o comportamientos que deberían demostrarse. Esto entraña una necesidad de atención a la uniformidad entre los criterios.

Una cuestión importante para poder diseñar evaluaciones prácticas que puedan aplicarse en una amplia gama de contextos del aula ha sido la determinación del concepto del tamaño adecuado para los criterios. Algunos investigadores (Kagan, 1990; Katz y Raths, 1985) han denominado a este reto el «Principio de Goldilocks». En la evaluación de los maestros, este principio significa que, si los criterios son demasiado extensos, es decir, demasiado vagos y generales, es difícil formular y aplicar equilibradamente estándares significativos; los evaluadores no pueden aplicar una serie uniforme de enjuiciamientos al proceso de evaluación. Por el contrario, si los criterios son demasiado estrechos (demasiado específicos), pueden enjuiciarse con gran uniformidad, pero no captarán la esencia de la buena enseñanza y pueden promover un enfoque fragmentado y prescriptivo de la enseñanza. Así pues, los criterios de evaluación, como el cocido de oso que probó Goldilocks, han de ser «perfectos» si se desea alcanzar las metas de una evaluación justa y una mejor práctica docente.

Como cuestión práctica, para encontrar el nivel de especificidad «perfecto» se

requieren muchas interacciones de trabajo de campo y análisis, tales como las que se ilustran en la Figura 2.

En los informes del trabajo de campo para las evaluaciones de la Práctica, hay numerosos casos de estas experiencias que llevan a los elaboradores a la conclusión, por ejemplo, de que lo que había sido un sólo criterio debería dividirse en dos criterios separados para ayudar a los evaluadores a comprender mejor cómo un determinado aspecto de la enseñanza se lleva a cabo realmente en el aula y ayudarles a reconocer la evidencia relacionada con este aspecto de la enseñanza cuando la observan.

Dicho trabajo de campo también permite a los evaluadores y elaboradores de las pruebas ver la forma en que los criterios guardan relación unos con otros en la práctica, y utilizar esta información como base para efectuar cambios en la forma en que los criterios se organizan, ordenan y describen. La organización y redacción de los criterios de forma que sean claros y lógicos, desde el punto de vista de quienes los van a utilizar, debería recibir alta prioridad en la labor de formulación.

Evaluadores y enjuiciamiento

Tal como se ha indicado antes, el enjuiciamiento desempeña una función importante en la elaboración de cualquier instrumento para la evaluación de los maestros. En las pruebas de opciones múltiples, el enjuiciamiento se ejerce al nivel de decidir cuestiones tales como qué someter a prueba, qué preguntas deberían presentarse, qué constituye una respuesta correcta y el número de preguntas que necesitan contestarse correctamente para que se considere que un candidato ha aprobado la prueba. Así pues, incluso en esta forma familiar de pruebas de «objetivos», el enjuiciamiento desempeña un papel importante, aunque no en la propia puntuación. Las evaluaciones del desempeño, por el contrario, incorporan enjuiciamiento en muchas formas que son similares, pero también en algunas que son diferentes. Al utilizar las evaluaciones del desempeño para evaluar a los maestros, es vital comprender el papel desempeñado por enjuiciamiento profesional de quienes los evalúan.

La especificación de los criterios de evaluación es una parte importante de la formulación de cualquier prueba de desempeño, pero el éxito de esta actividad en su conjunto sólo puede evaluarse a la luz de la capacidad de los evaluadores de utilizar estos criterios a fin de llegar a conclusiones técnica y profesionalmente defensibles.

Como en las pruebas tradicionales de opciones múltiples, en las que la gran preponderancia del enjuiciamiento profesional entra en juego durante la etapa previa a la administración de la formulación de la prueba, el enjuiciamiento profesional con las evaluaciones del desempeño se requiere tanto en las etapas de formulación como de utilización de la prueba. La calidad del enjuiciamiento profesional surte efecto sobre muchos aspectos importantes de la validez de la prueba, incluyendo, aunque sin limitarse a ello, la imparcialidad, complejidad

cognoscitiva y representación de los elementos. También guarda relación con las preocupaciones de generalización, aun cuando éstas no sean las mismas para una evaluación del desempeño natural en el aula que para la evaluación basada en la simulación.

En la evaluación del desempeño en el aula, la preocupación con la generalización gira en torno a otros eventos docentes, no a otros aspectos de la enseñanza. En este sentido, los eventos docentes son análogos a los ejercicios o tareas en otras clases de evaluaciones del desempeño. Un ejemplo de esto puede encontrarse en la Práctica III, Evaluaciones del Desempeño en el Aula (véase el Cuadro 1 donde se da una descripción de los criterios para evaluación en la Serie de Práctica).

En el modelo de la Práctica, la "puntuación" de las "tareas" se mantiene constante a través de los criterios y sus estándares de puntuación conexos. Al evaluar el desempeño docente en vivo, la variabilidad entre las "tareas" es un fenómeno natural y aceptable y, por tanto, se espera que las inferencias basadas en una serie dada de eventos docentes sean generales para un universo intrínsecamente variable de eventos docentes que define el concepto. Por tanto, la generalización entre las tareas no es problemática en el mismo sentido que cuando las tareas se consideran como elementos parciales o indirectos del concepto.

Tal como se indicó anteriormente, los criterios de la Práctica III, Evaluaciones del Desempeño en el Aula, tienen por fin ser interpretados como aspectos interrelacionados de un desempeño complejo, no como entidades funcionalmente independientes. Como tales, no trataríamos de generalizar de un aspecto de la enseñanza a otro como prueba de la validez sino, más bien, investigar los patrones de calificaciones dadas a través de distintas ocasiones y dentro de una ocasión dada por dos o más evaluadores (suponiendo que se espere que las ocasiones sean muy variables, en relación con la variabilidad del asesor o entre los asesores).

Los criterios de evaluación no son independientes, ya que constituyen aspectos de la enseñanza y no comportamientos particulares. Han de ser interpretados a la luz del contexto real en el aula, que incluye tanto a los alumnos como a la materia que se está enseñando. Los criterios sirven de guía para estructurar los enjuiciamientos de los evaluadores, asegurando que un marco común de referencia en vez de la preferencia personal constituya la base para las conclusiones de los evaluadores y las calificaciones. Por tanto, los enjuiciamientos profesionales de los evaluadores son la piedra angular del defendimiento de las puntuaciones del maestro que comienza a ejercer su profesión.

Utilizando los métodos que acabamos de describir, los evaluadores recopilan y organizan datos que inciden en cada uno de los criterios; efectúan enjuiciamientos críticos acerca de la importancia de la evidencia y su pertinencia para determinados criterios; luego llegan a una conclusión acerca del nivel de desempeño del maestro principiante en cada criterio basándose en esta evidencia y su interpretación de la misma. Los evaluadores documentan estos enjuiciamientos citando evidencia

concreta y vinculándola con una escala de calificaciones que describe niveles cada vez más suficientes de desempeño con respecto a cada uno de los criterios. La legitimidad del proceso de evaluación se basa por tanto en la calidad de este argumento (enjuiciamiento estructurado, documentado, profesional) más que en una ausencia pretendida de toma de decisiones humana (objetividad). Mediante estudios especiales (tales como las comparaciones por dos evaluadores), el trabajo de campo en distintos marcos docentes y el uso operativo, puede hallarse que los métodos de recopilación de datos distintos de los que ahora utilizamos producen mejor mención—es decir, enjuiciamientos exactos o detallados de los criterios, mejor documentación y efectos más positivos sobre el sistema de que forma parte la evaluación. Al determinar el valor y la validez de las evaluaciones, no obstante, los propios métodos de recopilación de datos están claramente subordinados a la calidad de los criterios y los enjuiciamientos y a la aplicación sistemática por los evaluadores de los procedimientos de evaluación.

CONCLUSIÓN

La evaluación de los maestros es un elemento clave de todo esfuerzo por mejorar la educación. Los sistemas bien concebidos y ejecutados para evaluación de los maestros son una parte integral del sistema educativo y coadyuvan al logro de sus metas principales. Los sistemas mal concebidos y ejecutados pueden tener un efecto deletéreo sobre el sistema mismo al que pretenden mejorar. Así pues: (1) una atención cuidadosa al diseño de las evaluaciones de los maestros, (2) seguimiento del cumplimiento con los planes durante la etapa de ejecución y (3) evaluación sistemática de todo el sistema de evaluación y sus consecuencias a través de su operación, son tres factores esenciales para obtener los beneficios previstos de las evaluaciones de los maestros.

Los educadores han tenido históricamente que luchar con la tensión entre las preocupaciones del equilibrio, la excelencia y la eficiencia que compiten entre sí. En distintos momentos, por razones comprensibles, el equilibrio ha cambiado a la preocupación acerca de uno de estos elementos a expensas de los otros. A fin de atender óptimamente las necesidades de la educación y de todos aquellos a quienes ésta afecta, no obstante, necesitamos esforzarnos por ir más allá de nuestras propias preocupaciones diarias e individuales, y mantener nuestra mirada concentrada en el cuadro más grande a fin de obtener el mejor equilibrio posible entre estas tres preocupaciones urgentísimas.

Cuadro 1: Criterios de evaluación para la Serie de Práctica: Evaluaciones del aula para maestros principiantes

---

Campo A: Organizar el conocimiento del contenido para el aprendizaje de los alumnos. El conocimiento del contenido que se ha de impartir sustenta todos los aspectos de la buena instrucción. El campo A se concentra en cómo los maestros utilizan su comprensión de los alumnos y las asignaturas para decidir acerca de las metas de

aprendizaje; diseñar o seleccionar actividades y material de instrucción apropiados; ordenar la instrucción de forma que ayude a los alumnos a alcanzar metas curriculares a corto y largo plazos; y concebir o seleccionar estrategias de evaluación informativa. Todos estos procesos, comenzando con las metas de aprendizaje, deben alinearse entre sí y, debido a las distintas necesidades representadas en toda aula, cada uno de los procesos mencionados debe llevarse a cabo de forma que responda a la serie de conocimientos y experiencias que los estudiantes traen al aula. Por consiguiente, el conocimiento de información pertinente acerca de los propios alumnos es parte integrante de este campo.

El campo A se refiere a cómo piensa el maestro acerca del contenido que va a enseñar. Este pensamiento es evidente en la forma en que el maestro organiza la instrucción en beneficio de sus alumnos. Las principales fuentes de evidencia en este campo son el Perfil del Aula, el Perfil de la Instrucción y la Entrevista Previa a la Observación. La observación del aula puede contribuir también a evaluar el desempeño en este aspecto.

Criterios de evaluación para el campo A:

A1:Familiarizarse con los aspectos relevantes de las experiencias y los conocimientos básicos de los alumnos.
A2:Formular metas claras de aprendizaje de las lecciones que sean apropiadas para los alumnos.
A3:Demostrar una comprensión de los vínculos entre el contenido que se aprendió con anterioridad, el contenido actual y el que queda por aprender en el futuro.
A4:Crear o seleccionar métodos de aprendizaje, actividades de aprendizaje y material u otras fuentes de instrucción que sean apropiados para los alumnos y que armonicen con las metas de la lección.
A5:Crear o seleccionar estrategias de evaluación que sean apropiadas para los alumnos y que armonicen con las metas de la lección.

Campo B: Crear un entorno para el aprendizaje del alumno. El campo B se relaciona con los componentes sociales y emocionales del aprendizaje como requisitos y contexto del rendimiento académico. Así pues, la mayor parte de los criterios de este campo se concentran en las interacciones humanas en el aula, en los vínculos entre maestros y alumnos y entre los alumnos. El campo B aborda las cuestiones relacionadas con la imparcialidad y la concordia, con ayudar a los alumnos a convencerse de que pueden aprender y pueden responder a los retos de establecer y mantener normas constructivas de comportamiento en el aula. También incluye el "entorno" de aprendizaje en el sentido más literal: el marco físico en el que tienen lugar la enseñanza y el aprendizaje.

Un entorno de aprendizaje que proporciona seguridad tanto emotiva como física a los alumnos puede propiciar una amplia gama de experiencias de enseñanza y aprendizaje. Los maestros deben estar en condiciones de utilizar el conocimiento que tienen de sus alumnos para interpretar con exactitud el comportamiento de ellos y responder de manera apropiada y ofreciendo apoyo. Al hacerlo, sus interacciones con los alumnos promueven continuamente el sentido de autoestima en ellos. Además, los esfuerzos de los alumnos por crear un sentido de comunidad en el aula con estándares claros no deberían ser nunca arbitrarios; todos los estándares de comportamiento e interacciones entre el maestro y los alumnos deberían fundamentarse en un sentido de respeto hacia todos los miembros de la comunidad en el aula.

Las pruebas de los criterios del campo B provienen principalmente de la observación del aula; las pruebas de apoyo pueden provenir de entrevistas tanto anteriores como posteriores a la observación. El Perfil del Aula proporciona información de contexto pertinente para los criterios que encierra este campo.

Criterios de evaluación del campo B:

B1:Crear un clima que promueva la imparcialidad.
B2:Establecer y mantener una relación con los alumnos.
B3:Comunicar expectativas de aprendizaje interesante para cada alumno.
B4:Establecer y mantener estándares uniformes de comportamiento en el aula.
B5:Hacer el entorno físico lo más seguro y conducente que sea posible al aprendizaje.

Campo C: Enseñanza para el aprendizaje de los alumnos. Este campo se concentra en el acto de enseñar y en su meta general: ayudar a los alumnos a conectarse con el contenido. Tal como se utiliza aquí, el "contenido" se refiere a la materia de una disciplina y puede incluir conocimientos, aptitudes, percepciones y valores en cualquier campo: cognoscitivo, social, artístico, físico y demás. Los maestros dirigen a los alumnos en el proceso de establecer conexiones individuales con el contenido, haciendo así "encajar" bien el contenido en el marco de los conocimientos, intereses, aptitudes, antecedentes culturales y antecedentes personales de los alumnos. Al mismo tiempo, los maestros deberían ayudar a los alumnos a sobrepasar los límites de sus conocimientos o comprensión actuales. Los maestros supervisan el aprendizaje, asegurándose de que los alumnos asimilen con exactitud la información y que comprendan y puedan poner en práctica lo que han aprendido. Los maestros deben estar seguros también de que los alumnos comprendan lo que se espera de ellos en cuanto a procedimientos durante la lección y que el tiempo de clase se utilice provechosamente.

Gran parte de esta evidencia para el desempeño del maestro con respecto a estos criterios provendrá de la observación del aula. Ésta puede aumentarse o iluminarse con la evidencia proveniente de entrevistas anteriores o posteriores a la observación, del Perfil de la Instrucción y del Perfil de la Clase.

Criterios de evaluación del campo C:

C1:Aclarar a los alumnos las metas de aprendizaje y los procedimientos de instrucción.
C2:Hacer el contenido comprensible para los alumnos.
C3:Alentar a los alumnos a ampliar su capacidad de reflexión.
C4:Vigilar la comprensión del contenido por parte de los alumnos mediante una gama de medios, proporcionando intercambio de información a los alumnos con el fin de apoyar el aprendizaje y adaptando las actividades de aprendizaje según lo exija la situación.
C5:Utilizar eficazmente el tiempo dedicado a la instrucción.

Campo D: Profesionalismo de los maestros. Los maestros han de poder evaluar su propia eficacia en la instrucción a fin de planificar lecciones futuras específicas para determinadas clases y mejorar su enseñanza a través del tiempo. Deberían poder hablar del grado en que los distintos aspectos de una lección tuvieron éxito en términos de métodos de instrucción, respuestas de los alumnos y resultados del aprendizaje. Los maestros deberían poder explicar cómo procederán para trabajar en pro del aprendizaje de todos los alumnos. Las responsabilidades profesionales de todos los maestros, incluidos los maestros

principiantes, también entrañan compartir información adecuada con otros profesionales y con las familias de forma que apoyen el aprendizaje de diversas poblaciones de alumnos.

Criterios de evaluación del campo D:

D1:Reflexionar sobre el grado en que se alcanzaron las metas de aprendizaje.
D2:Demostrar un sentido de eficacia.
D3:Crear relaciones profesionales con colegas para compartir conocimientos acerca de la enseñanza y coordinar las actividades de aprendizaje de los alumnos.
D4:Comunicarse con los padres de familia o tutores acerca del aprendizaje del alumno.

NOTAS

[1] En el original inglés se utiliza el término *evaluation* a lo largo de todo este documento para indicar una gama de actividades emprendidas con el objeto de recopilar información para fines de toma de decisiones. El término se utiliza sin referencia a distinciones entre dichos fines e incluye, pero sin limitarse a ello, debates de técnicas específicas de evaluación. El término *assessment* se utiliza para referirse a estas técnicas específicas de recopilación de datos. En la actualidad, el uso de estos dos términos en los Estados Unidos es algo controvertido, refiriéndose el primero principalmente a actividades que realizan con maestros practicantes quienes los emplean, con el fin de tomar decisiones de personal aisladas.

[2] El término *puntuación límite* se utiliza en esta sección para referirse a todo estándar que indique un nivel de rendimiento. El uso de este término se aplica a cualquier clase de evaluación en la que las personas evaluadas van a ser divididas en grupos para fines de toma de decisiones.

REFERENCIAS

American Educational Association, American Psychological Association, y National Council on Measurement in Education. (1985). *Standards for Educational and Psychological Testing*. Washington, DC: American Psychological Association (APA).

Andrews, H. A. (1985). *Evaluating for Excellence: Addressing the Need for Responsible and Effective Faculty Evaluation*. Stillwater, OK: New Forums Press.

Banks, C. A. M. (1997). The challenges of national standards in a multicultural society. *Education Horizons*, 75, 3: 126-132.

Berk, R. A. (1988). Fifty reasons why student achievement gain does not mean teacher effectiveness. *Journal of Personnel Evaluation in Education,* 1: 345-363.

Bridges, E. (1986). *The Incompetent Teacher: The challenge and the response*. Filadelfia: Falmer Press.

Burry, J. A., Chissom, B. S., y Shaw, D. G. (1990, March). *Validity and Reliability of Classroom Observations: A Paradox*. Paper presented at a meeting of the National

Council on Measurement in Education, Boston.

Claxton, C., Murrell, P. H., y Porter, M. (1987). Outcomes Assessment. *AGB Reports, 29*, 5: 32-35.

Cole, N. S., y Moss, P. A. (1989). Bias in Test Use. En: R. L. Linn (Ed.), *Educational Measurement* (3rd ed., págs. 201-219). Nueva York: Macmillan.

Darling-Hammond, L. (1986). A proposal for evaluating in the teaching profession. *Elementary School Journal,* 86: 531-569.

Duke, D., y Stiggins, R. J. (1990). Beyond Minimum Competence: Evaluation for Professional Development. En: J. Millman y L. Darling-Hammond (Eds.), *The New Handbook of Teacher Evaluation: Assessing Elementary and Secondary School Teachers* (págs. 116-132). Newbury Park, CA: Sage.

Dwyer, C. A. (1993b). "Innovation and Reform: Examples from teacher assessment." En: R. E. Bennett y W. C. Ward (Eds.), *Construction Versus Choice in Cognitive Measurement* (págs. 265-289). Hillsdale, NJ: Lawrence Erlbaum.

Dwyer, C. A. (1993c). Teaching and diversity: Meeting the challenges for innovative teacher assessments. *Journal of Teacher Education,* 44(2), 119-129.

Dwyer, C. A. (1994). Criteria for Performance-Based Teacher Assessments: Validity, Standards, and Issues. *Journal of Personnel Evaluation in Education, 8*, 135-150.

Dwyer, C. A. (1996). Cut-scores and Testing: Statistics, Judgment, Truth, and Error. *Psychological Assessment, 8* (4).

Dwyer, C. A., y Ramsey, P. A. (1995). "Equity Issues in Teacher Assessment." En: M. T. Nettles and A. L. Nettles (Eds.), *Equity and Excellence in Educational Testing and Assessment*. Boston: Kluwer Academic.

Dwyer, C. A. y Stufflebeam, D. (1996). Teacher Evaluation. En: D. Berliner y R. Calfee (Eds.), *Handbook of Educational Psychology* (págs. 765-786). Nueva York: Macmillan.

Dwyer, C. A. y Villegas, A. M. (1993). *Guiding Conceptions and Assessments Principles for The Praxis Series: Professional Assessments for Beginning Teachers* (RR 93-17). Princeton, NJ: Educational Testing Service.

Evertson, C., y Green, J. (1986). Observation as Method and Inquiry. En: M. Wittrock (Ed.), *Handbook of Research on Teaching* (3rd ed., págs. 162-213). Nueva York: Macmillan.

Hunter, M. (1988). Effecting a Reconciliation Between Supervision and Evaluation. *Journal of Personnel Evaluation in Education, 1*: 275-279.

Garcia, P. A. (1985). *A Study of Teacher Competency Testing and Tests Validity with Implications for Minorities: Final Report.* (NIE Grant No. NIE-G-85-0004). Edinburgh, TX: Pan American University.

Glass, G. V. (1990). Using Student Test Scores to Evaluate Teachers. En: J. Millman y L. Darling-Hammond (Eds.), *The New Handbook of Teacher Evaluation* (págs. 191-215). Newbury Park, CA: Sage.

Haney, W., Madaus, G., y Kreitzer, A. (1987). Charms Talismanic: Testing Teachers for the Improvement of American Education. En: E. Z. Rothkopf (Ed.), *Review of Research in Education* (Vol. 14, págs. 169-238). Washington, DC: American Educational Research Association.

Hood, S., y Parker, L. (1991). Minorities, Teacher Testing, and Recent U.S. Su-

preme Court Holdings: A Regressive Step. *Teachers College Record,* 92: 603-618.
Joint Committee on Standards for Educational Evaluation. (1988). *The Personnel Evaluation Standards*. Newbury park, CA: Sage.
Kagan, D. M. (1990). Ways of Evaluating Teacher Cognition: Inferences Concerning the Goldilocks Principle. *Review of Educational Research,* 60, 419-469.
Lawshe, C. H. (1975). A Quantitative Approach to Content Validity. *Personnel Psychology, 28*: 563-575.
Linn, R. L., Baker, E. L., y Dunbar, S. B. (1991). Complex, Performance-Based Assessment: Expectations and Validation Criteria. *Educational Researcher, 20* (8): 15-21.
Madaus, G. F. (1990). Legal and Professional Issues in Teacher Certification Testing: A Psychometric Snark Hunt. En: J. V. Mitchell, S. L. Wise, and B. S. Plake (Eds.), *Assessment of Teaching: Purposes, Practices, and Implications for the Profession* (págs. 209-259). Hillsdale, NJ: Lawrence Erlbaum.
Madaus, G. y Mehrens, W. A. (1990). Conventional Tests for Licensure. En: J. Millman y L. Darling-Hammond (Eds.), *The New Handbook of Teacher Evaluation: Assessing Elementary and Secondary School Teachers* (págs. 257-277). Newbury Park, CA: Sage.
Madaus, G. F., y Pullin, D. (1987, September). Teacher Certification Tests: Do They Really Measure What We Need to Know? *Phi Delta Kappan*, 69(1), 31-38.
Mehrens, W. A. (1987). Validity Issues in Teacher Licensure Tests. *Journal of Personnel Evaluation in Education, 1:* 195-229.
Messick, S. (1989). Validity. En: R. L. Linn (Ed.), *Educational Measurement* (3rd ed., págs. 13-103). Nueva York: Macmillan.
Miller, M. D., y Legg, S. M. (1993). Alternative Assessment in a High-Stakes Environment. *Educational Measurement: Issues and Practice, 12* (2): 9-15.
Millman, J. (1981a). *Handbook of Teacher Evaluation*. Beverly Hills, CA: Sage.
Millman, J. (1981b). Student Achievement as a Measure of Teacher Competence. En: J. Millman (Ed.) *Handbook of Teacher Evaluation* (págs. 146-166). Beverly Hills, CA: Sage.
Milner, J. O. (1991). Suppositional Style and Teacher Evaluation. *Phi Delta Kappan,* 72: 464-467.
Moss, P. A. (1992). Shifting Conceptions of Validity in Educational Measurement: Implications for Performance Assessment. *Review of Educational Research, 62*: 229-258.
Murnane, R. J., y Schwinden, M. (1989). Race, Gender, and Opportunity: Supply and Demand for New Teachers in North Carolina, 1975-1985. *Educational Evaluation and Policy Analysis, 11*: 93-108.
National Board for Professional Teaching Standards (1991). *Toward High and Rigorous Standards for the Teaching Profession* (3rd ed.). Detroit, MI: Autor.
National Council of Teachers of Mathematics. (1989, March). *Curriculum and Evaluation Standards for School Mathematics*. Reston, VA: Autor.
Quellmalz, E. S. (1991). Developing Criteria for Performance Assessments: The Missing Link. *Applied Measurement in Education, 4*: 319-331.
Rebell, M. A. (1990). Legal Issues Concerning Teacher Evaluation. En: J. Millman and L. Darling-Hammond (Eds.), *The New Handbook of Teacher Evaluation:*

Assessing Elementary and Secondary School Teachers (págs. 337-355). Newbury Park, CA: Sage.

Redfern, G. B. (1960). *How to Appraise Teaching Performance*. Columbus, OH: School Management Institute.

Redfern, G. B. (1980). *Evaluating Teachers and Administrators: A Performance Objectives Approach*. Boulder, CO: Westview Press.

Scriven, M. (1987). Validity in Personnel Evaluation. *Journal of Personnel Evaluation in Education,* 1: 9-23.

Scriven, M. (1988a). Duties-Based Teacher Evaluation. *Journal of Personnel Evaluation in Education,* 1: 319-334.

Scriven, M. (1988b). Evaluating Teachers as Professionals: The Duties-Based Approach. En: S. J. Stanley y W. J. Popham (Eds.), *Teacher Evaluation: Six Prescriptions for Success* (págs. 110-142). Alexandria, VA: Association for Supervision and Curriculum Development.

Sedlak, M., y Schlossman, S. (1986). *Who Will Teach? Historical Perspectives on the Changing Appeal of Teaching as a Profession*. (Report No. 3472--CSTP). Santa Monica, CA: RAND Center for the Study of the Teaching Profession.

Shalock, D., y Shalock, M. (1990, September). Extending Teacher Assessment Beyond Knowledge and Skills: An Emerging Focus on Teacher Accomplishments. *Issues and Practices in Performance Assessment,* 2: 81-126.

Shimberg, B. (1985). Testing for Licensure and Certification. *American Psychologist, 36*: 1138-1146.

Shulman, L. S. (1986) Paradigms and Research Programs in the Study of Teaching: A Contemporary Perspective. En: M. C. Wittrock (Ed.), *Handbook of Research on Teaching* (3rd ed., págs. 3-36). Nueva York: Macmillan.

Shulman, L. S. (1987). Knowledge and Teaching: Foundations of the New Reform. *Harvard Educational Review, 57*: 1-22.

Shulman, L. S. (1988). A Uion of Insufficiencies: Strategies for Teacher Assessment in a Period of Educational Reform. *Educational Leadership,* 46: 36-41.

Shulman, L. S. (1988a). A Union of Insufficiencies: Strategies for Teacher Assessment in a Period of Educational Reform. *Educational Leadership,* 46: 36-41.

Shulman, L. S. (1988b). The Paradox of Teacher Assessment. En: *New Directions for Teacher Assessment. Proceedings of the 1988 ETS Invitational Conference*. Princeton, NJ: Educational Testing Service.

Shulman, L. S. (1989). The Paradox of Teacher Assessment. En: *New Directions for Teacher Assessment: Proceedings of the 1988 ETS Invitational Conference* (págs. 13-27). Princeton, NJ: Educational Testing Service.

Stodolsky, S. S. (1988). *The Subject Matters.* Chicago: University of Chicago Press.

Streifer, P., y Iwanicki, E. (1987). The Validation of Beginning Teacher Competencies in Connecticut. *Journal of Personnel Evaluation in Education,* 1: 33-55.

Webster, W. J., Mendro, R. L., y Almaguer, T. O. (1993). *Effectiveness Indices: The Major Component of an Equitable Accountability System.* Estudio presentado en la reunión anual de la American Educational Research Association, Atlanta, GA.

Wise, A. E., Darling-Hammond, L., McLaughlin, M. W., y Bernstein, H. T. (1984a). *Case Studies for Teacher Evaluation: A Study of Effective Practices.* Santa Monica, CA: Rand Corporation.

Wise, A. E., Darling-Hammond, L., McLaughlin, M. W., y Bernstein, H. T. (1984b). *Teacher Evaluation: A Study of Effective Practices.* Santa Monica, CA: Rand Corporation.

CAPÍTULO 8

EVALUACIÓN DE LA ACCIÓN DOCENTE EN LATINOAMÉRICA

Francisco Álvarez Martín, con la colaboración de María José Álvarez y Paula Vergara[1]

El capítulo anterior presentó un panorama amplio de la evaluación de los maestros. Este capítulo trata esta misma problemática, pero desde la perspectiva de la práctica en una muestra de cinco países de América Latina. El documento identifica algunos modelos de evaluación propuestos en la región por investigadores, o ensayados en escala reducida en algunas innovaciones educativas. Además, discute los problemas políticos, técnicos y de recursos con los cuales se enfrenta una iniciativa encaminada a establecer sistemas nacionales de evaluación de la profesión docente.

INTRODUCCIÓN

El objetivo de este trabajo es presentar, a partir de la información existente en la Red Latinoamericana de Información y Documentación Educativa (REDUC), el panorama general de las prácticas de evaluación del desempeño docente en América Latina y hacer sugerencias de políticas al respecto.

Bajo el descriptor "evaluación del docente" se han encontrado sólo 78 títulos (con desigual pertinencia) en la colección de resúmenes analíticos publicada por REDUC, entre un total de 7.925 publicaciones sobre educación resumidas hasta el segundo semestre de 1996. En los últimos diez años se ha publicado casi la totalidad de estos trabajos. Además, también cabe destacar que las instituciones patrocinantes de estos estudios son en el 50 por ciento de los casos instituciones privadas, 39 por ciento universidades, 6 por ciento ministerios de educación y 5 por ciento organismos internacionales.

La mayoría de estos trabajos (53 por ciento) abordan el tema en forma genérica sin referencia a ningún país o región; otro grupo considerable (27 por ciento) trata el tema de la evaluación docente dentro de la perspectiva y práctica universitarias, enfocándolo tanto desde el punto de vista teórico como también desde el análisis de

las experiencias prácticas. El resto de los trabajos se refiere al tema de evaluación del docente del nivel secundario o de la escuela técnica profesional (13 por ciento), apareciendo como escasos los referidos al tema dentro del nivel de la enseñanza básica (5 por ciento).

La evaluación de los docentes no ha sido un tema prioritario en América Latina, lo cual no significa que no haya existido una práctica y legislación o normativa al respecto. La evaluación es algo que siempre se ha hecho y se hace de manera permanente en las instituciones dedicadas a la educación. Tanto los directores de centros docentes, como también los alumnos y las mismas familias de éstos, aunque sea con mecanismos no formales, están evaluando el comportamiento pedagógico de los maestros y las maestras. Sin embargo, los criterios y las perspectivas de evaluación difieren mucho unos de otros. En lo que sí parece darse una constante, es en que cada vez que se propone hacer una evaluación con carácter sistemático, los docentes la consideran como una amenaza.

Las reformas educativas que se ponen en marcha en la región a partir de los años sesenta tienen como foco central la cobertura y están influenciadas por una posición tecnológica la cual enfatiza que la eficacia de la enseñanza está más ligada a la calidad o al tecnicismo con que se ejecutan ciertas tareas específicas que a las características o cualidades del profesor. De estos enfoques se lograron importantes avances que se manifiestan en la universalización de la educación y en un incremento de las tasas de asistencia anual, pero no así en las dinámicas educativas que se dan al interior de las aulas escolares.

En la década de los años ochenta, como fruto de los estudios preparatorios para la Conferencia Mundial de Educación para Todos celebrada en Jomtien, Tailandia, el foco de las políticas educativas se vuelve hacia la efectividad y calidad de la educación. Namo de Mello (1982) realiza estudios en los que se reconoce que el crecimiento cuantitativo del sistema ha impuesto, de hecho, un importante cambio cualitativo en la dinámica escolar. A partir de dichas investigaciones el tema comienza a aparecer con mayor fuerza en la literatura educativa. Es el caso, por ejemplo, de los estudios que Ezpeleta (1989) y Tovar (1989) hacen sobre la situación de los maestros en la Argentina, el Perú y Bolivia, así como las investigaciones de Arancibia (1994) centradas en la efectividad escolar en las que predominan el estudio de los factores que inciden en la calidad de la enseñanza.

En los últimos tiempos el tema ha cobrado especial relevancia por asociarse a lo que aparece como foco central de las reformas educativas que se ejecutarán en la región. Esto aparece con mucha claridad en el proyecto presentado por la Universidad Nacional de Colombia al Ministerio de Educación Nacional para la puesta en marcha de un Sistema Nacional de Evaluación (1996). En dicho documento se enfatiza que las nuevas concepciones y los propósitos educativos requieren una evaluación de los maestros que los motive y prepare para asumir responsablemente el impacto de las reformas en la sociedad.

De políticas dirigidas a asegurar la mayor cobertura del sistema educativo, se ha pasado a políticas orientadas a garantizar una mayor calidad en la educación. De ahí que surja como preocupación importante lo que se enseña y lo que se aprende. Y si bien se reconoce que son muchos los factores que intervienen en el aprendizaje de los niños, la actuación de los que están a cargo del aprendizaje, es un factor clave al que hay que dedicar mayor atención.

Sin embargo, no se puede decir que haya consenso respecto a la necesidad no sólo de investigar sobre el tema, sino también de legislar y promover acciones para el desarrollo de sistemas de evaluación. Mientras que algunos especialistas como Connelly (1989) afirman que la evaluación de profesores consume bastante tiempo, cuesta demasiado y tiene muy pocos beneficios demostrables para justificar su continuación, Ahumada (1992) sostiene que la evaluación de la eficiencia docente constituye uno de los problemas más interesantes y tal vez menos estudiados en los desarrollos y procesos educativos.

Según se mantenga una posición u otra se admitirá que tanto la falta de modelos y metodología como la ignorancia sobre las aplicaciones invitan a improvisaciones que reciben rechazo por parte de los afectados como señala Ahumada (1992). O bien, se insistirá en que hay que dar prioridad a la educación de los profesores durante su etapa de formación, antes que a los docentes en servicio como sostiene Connelly (1989).

El presente trabajo está estructurado en tres partes: En la primera parte se recogen los conocimientos que nos aportan las investigaciones sobre el tema. Estos conocimientos se refieren a lo siguiente: a) el concepto mismo de "evaluación del desempeño docente"; b) los criterios para la evaluación; c) los aportes de las innovaciones educativas; d) los modelos de evaluación que surgen desde la investigación y e) la evaluación del docente desde la perspectiva de los gremios. En la segunda parte se estudian los casos de algunos países considerando la normativa existente sobre este tema y la práctica que hay de la misma. En la tercera parte se hacen recomendaciones sobre investigación y política a partir de los estudios y casos revisados.

¿QUÉ SABEMOS SOBRE LA EVALUACIÓN DE LOS DOCENTES?

Concepto de "evaluación del desempeño docente"

Las experiencias y los sistemas de evaluación docente que se viven en la región surgen de la adaptación de los Sistemas de Evaluación de Empresa. Se quiso imponer para la institución escolar la tendencia conocida en la administración de empresa como "Desarrollo Organizacional" que busca aumentar la eficacia de las instituciones. Sin embargo, aunque la escuela quiera ser vista como empresa siempre va a tener rasgos que la distinguen de cualquier otra empresa, tales como sus finalidades y "productos". Tal vez de ahí surge la mayoría de preguntas, incoherencias y vacíos que ocurren con la evaluación del desempeño docente.

Los estudios de Ahumada (1992) intentan explicar la crítica y el reclamo de los docentes frente a cualquier intento evaluativo, ya que la mayoría de las veces los resultados de este proceso suelen ser usados con propósitos que hacen que la evaluación se convierta en un atentado a la estabilidad laboral en lugar de ser una ayuda para mejorar la enseñanza.

Villa Sánchez (1985) mantiene que no siempre que se habla de evaluación docente se está hablando de lo mismo. El concepto de evaluación docente varía según su finalidad sea el ascenso o la mejora de la tarea docente. En la misma línea, Connelly (1989) señala cómo se usan los formularios para la calificación de los maestros registrando más tareas superficiales (rutina administrativa) que tareas de desempeño; por su parte, Carranza (1992) subraya que los resultados de los mismos, en vez de estar al servicio del profesor para ayudar a mejorar su trabajo, se usan como medio coercitivo. En este sentido de una búsqueda de mayor coherencia, es que hay que entender la propuesta de Villa Sánchez (1985) cuando dice que cada sistema de evaluación del profesor debe poseer una definición de la tarea docente y un procedimiento para evaluar.

Además, algunos opinan que la diferencia entre los modelos existentes surge también por la diversidad de instrumentos que es necesario poner en juego. Ya que si bien hay un cierto consenso en el que el principal objetivo de la evaluación docente es "determinar las cualidades profesionales, la preparación y el rendimiento del educador" según Carranza (1992) o "la medición de las conductas del maestro" según Garro (1988), no hay instrumento que haya sido técnicamente elaborado y que pueda dar información suficiente sobre la calidad del docente o la docencia misma. Por el contrario, como señala Ahumada (1992), es necesario recurrir a varios instrumentos que aporten una información procedente de diferentes fuentes, momentos o procedimientos. Desde esta perspectiva hay que entender la propuesta que hacen Ordoñez y otros (1996) cuando señalan que la evaluación de los docentes debe pensarse como un fundamento para la elaboración de planes y programas de mejora y que requiere de diversas miradas.

No obstante, hay que reconocer que la evaluación del desempeño docente es un proceso en desarrollo que necesita de investigación continua, admitiendo que lo hecho hasta ahora ha permitido determinar algunos indicadores generales y específicos no sujetos a duda y relacionados con lo que para la mayoría corresponde a un docente exitoso. Este es ciertamente otro aporte importante de la investigación sobre el tema.

Criterios que se usan para la evaluación

No es fácil determinar con precisión y claridad qué se quiere evaluar cuando se habla de desempeño docente. Por un lado, es difícil establecer un consenso respecto al perfil del maestro eficaz. Por otro, se reconoce que en el acto de enseñar, que es sumamente complejo, intervienen muchos otros factores que la sola acción del

docente. De ahí que se considere como un aporte interesante el que hace la investigación respecto a criterios que hay que tener en cuenta para la evaluación del desempeño docente. Estos criterios surgen teniendo en cuenta lo que se hace en la práctica y cuestionando los efectos y resultados de la misma.

En el estudio de Avalos y Haddad (1981) sobre la efectividad de los maestros se intenta identificar cuáles son las variables relacionadas con el maestro que pueden estar asociadas con cambios en la situación docente o en los estudiantes, las escuelas y la comunidad. En este estudio, dichas variables aparecen divididas en dos grandes categorías: características del maestro (sexo, edad, personalidad, nivel socioeconómico, conocimientos, destrezas, idioma en el que enseña, actitudes, experiencia, calificaciones y capacitación) y características del sistema escolar (localización de las escuelas, tipo de administración escolar, disciplina y nivel de enseñanza, recursos, sistema de exámenes, carga de trabajo, incentivos, salario y *estatus* social del maestro).

Según dicho estudio, la mayoría de las investigaciones consideran los antecedentes contextuales del maestro y sus actitudes y comportamientos en el aula, pero muy pocas relacionan estos factores del maestro con los resultados y cambios de actitud de los alumnos.

Otro enfoque se ha aplicado desde 1960, a saber, el Cuestionario Descriptivo del Clima Organizacional con el que se puede determinar la existencia de una realidad o "un clima organizacional" al medir conductas percibidas en los profesores y directivos docentes. Esta técnica que pone énfasis en los aspectos relacionales, tanto dentro como fuera del aula escolar, constituye un instrumento de retroalimentación que permite a los profesores y también a sus superiores determinar acciones para superar debilidades del comportamiento organizativo y, en consecuencia, mejorar el desempeño y la eficacia docente. Sin embargo, según Rossenfeld (1983) no es tan fácil de aplicar ya que requeriría cierta preparación previa por parte de los docentes y adaptaciones a la idiosincrasia de las escuelas de la región.

Un estudio que trata sobre la evaluación del desempeño del maestro en el proceso didáctico, realizado por Galo de Lara (1990) en Guatemala, se centra en el comportamiento de los docentes en el proceso de enseñanza-aprendizaje. Para ello elabora un instrumento de observación basado en un sistema de categorías, el cual puede modificarse en su forma para ser utilizado como instrumento de autoevaluación. Las categorías permiten establecer interrelaciones que mantienen la congruencia del instrumento y , al mismo tiempo, señalan los aspectos del desempeño docente que deben mejorarse. Este instrumento eminentemente cualitativo (cuadro 1), está dirigido a propiciar la reflexión y la toma de conciencia del maestro sobre sus comportamientos en la dirección del proceso didáctico, a fin de que pueda perfeccionar su desempeño.

Cuadro 1. Categorías y dimensiones para la observación del proceso didáctico

El estudio considera la evaluación como un PROCESO en el que intervienen las variables que constituyen los elementos integrantes de lo que ocurre en el aula:

- actividades de enseñanza-aprendizaje
- comportamiento del maestro y los alumnos
- interrelaciones e interacciones y tipo de clima o ambiente de la clase.

Establece también que un buen maestro es el que cumple con los elementos normativos que estructuran el modelo y que aparecen en el cuadro anteriormente presentado.

Justa Ezpeleta (1989), a propósito de una investigación sobre las condiciones de los maestros en la Argentina, señala que el sistema educativo ha pensado siempre en un "maestro abstracto" sin condicionamientos sociales ni aspiraciones o proyectos personales. Esto lleva a plantear que es imposible abordar el tema de la evaluación del desempeño docente sin tener en cuenta los referentes del contexto sociocultural en el que los maestros se desenvuelven. Tanto en el estudio de Ezpeleta como en otro similar de Tovar (1989) sobre los maestros en el Perú, se evidencia que las prácticas cotidianas escolares están afectadas por otras situaciones que rebasan las capacidades didácticas de los maestros. Es importante tener esto en cuenta al establecer los criterios para evaluar el desempeño de los docentes.

Los estudios de Arancibia (1987 y 1994) tratan de identificar las conductas de los profesores efectivos (entendiendo como tal es a aquéllos que tienen alumnos con un alto rendimiento escolar) y muestran que en el rendimiento académico de los alumnos no sólo influye lo que el profesor hace en el aula sino también lo que el profesor es como persona, con sus expectativas, creencias, sentimientos y satisfacciones.

"La búsqueda incesante de criterios, indicadores y procedimientos para evaluar la eficiencia del docente parece estar llegando a su fin"—expresa Ahumada (1992), quien hace un esfuerzo por mostrar el nivel de consenso al que se ha llegado en el campo de la investigación en distintos países y sugiere que ya estamos en época de utilizar los criterios concordantes y, a partir de ellos, elaborar instrumentos flexibles y congruentes con los propósitos evaluativos que se persigan. Basándose en el modelo bidimensional de Popham (1980), que considera las formas de obtención de información y los propósitos asociados a dichas fórmulas, Ahumada agrega una tercera dimensión que se refiere al momento en que se efectúa la evaluación y construye así un modelo tridimensional de selección de criterios-tipo (cuadro 2) que permiten la elaboración de instrumentos congruentes con los propósitos de la evaluación.

El estudio de Connelly (1989) es sumamente crítico respecto a las diferentes perspectivas y/o criterios desde los que se ha abordado la evaluación del desempeño docente, pues considera que no se ha tenido en cuenta a los mismos docentes.

Destaca que la evaluación entre colegas es una forma de educación en servicio dirigida fundamentalmente a la mejora de la instrucción. Uno de sus reparos es que en algunos países en vías de desarrollo, el nivel de capacidad dentro de una profesión no sería suficientemente alto para justificar tal sistema. Una alternativa sugerida por el autor es lo que él denomina una "práctica reflexiva supervisada". El autor entiende esta práctica como un proceso dirigido a la educación de profesores, más que un sistema dirigido al despido o promoción de los mismos. Este proceso es una lección de sentido común, de cómo los profesores pueden desarrollar sus puntos fuertes mediante consultas con supervisores y colegas con respecto a sus prácticas de enseñanza.

Cuadro 2. Criterios de elaboración de instrumentos evaluativos

1. Adecuada organización del programa y del curso. Se refiere a la presentación de los objetivos y contenidos determinados, a la estrategia metodológica y evaluativa y a la bibliografía recomendada.

2. Comunicación afectiva (motivación - manejo de materiales). Está relacionada con el proceso efectivo de instrucción en el que el estudiante logra mayor éxito como efecto de una correcta utilización de medios de apoyo. Estos últimos pueden ser medios escritos y también aspectos relacionados con la claridad del mensaje comunicativo.

3. Conocimiento y entusiasmo por la materia de enseñanza (dominio del tema). Se refiere al grado de dominio que el docente tiene de su disciplina y que se manifiesta en la capacidad de relacionar lo que enseña con problemáticas de los que aprenden y de su mundo.

4. Actitudes positivas hacia los estudiantes (grado de relación profesor-alumno). Es una variable importante en el proceso de intercomunicación continua y se manifiesta en la aceptación, por parte del profesor, de errores o puntos de vista distintos de sus estudiantes, en su accesibilidad dentro y fuera de la clase, en la estimulación permanente a la participación, respeto mutuo y disposición a recibir y responder consultas.

5. Flexibilidad didáctica. Se refiere a la adaptación de la metodología y los recursos existentes según las capacidades de los estudiantes; a la habilidad para distribuir tiempos, atención a grupos chicos y grandes, selección de actividades y manejo de materiales técnicos según los distintos niveles que se dan en el curso.

6. Ecuanimidad en las evaluaciones y asignación de calificaciones. Relativo a la selección de criterios que aseguran una evaluación justa e imparcial, así como al diseño, corrección y calificación de pruebas de acuerdo a los principios técnicos que aseguran su validez y confiabilidad.

La práctica reflexiva-supervisada es una versión de la "reflexión sobre la acción" de

Shön (1987), la cual es un proceso emprendido deliberadamente para revisar lo que se ha hecho y lo que se podría hacer mejor. En su forma más simple, ésta podría ocurrir en tres etapas, a saber:

- sesión de observación,
- sesión de conversación sobre lo que se registra en la observación, y
- conversación de seguimiento de la que pueden surgir nuevas observaciones.

Esta estructura tripartita puede ser ampliada y modificada de varias maneras. Su proceso de aplicación podría ser tan importante para la formación inicial de los docentes como para la formación de los que están en servicio.

La clarificación de criterios es un elemento importante que puede surgir de la investigación y también de la misma práctica ya que hay experiencias innovadoras que plantean formas y procedimientos. Uno de estos elementos nuevos tiene que ver con que el proceso mismo de evaluación docente debe hacerse con la participación de los interesados y no solamente a través de agentes externos. Al respecto, Ordóñez y otros (1996) plantea que la evaluación sólo puede incidir en la transformación de la educación si se realiza con la participación de los directos encargados de la cotidianeidad de la escuela.

Aporte de las innovaciones

Las experiencias exitosas que se han llevado a cabo en educación y que se pueden llamar «innovadoras» por la ejecución de estrategias que han influido cualitativamente en los aprendizajes de los niños, han puesto en evidencia la importancia de establecer mecanismos de participación para que los docentes revisen su acción, criticándola e introduciendo en la misma modificaciones para transformarla cualitativamente. Este trabajo se limita sólo a presentarlas desde la óptica del componente de evaluación de la acción docente.

Escuela Nueva de Colombia

La evaluación del docente en la Escuela Nueva (EN)—programa escolar novedoso desarrollado en Colombia (cfr. Schiefelbein, 1993)— está concebida como una mirada permanente sobre la acción educativa con el propósito de tomar decisiones y transformar cualitativamente tal acción. En dicha mirada participan fundamentalmente los maestros y también otros agentes como los supervisores y coordinadores de microcentros; pero no están ausentes de esta mirada otros agentes externos a la escuela como los expertos e investigadores que estudian, fundamentalmente, el rendimiento y los factores que lo determinan.

a) Los maestros en sus microcentros disponen, de manera regular y sistemática, de espacio donde interactuar e intercambiar sus experiencias pedagógicas. Ello les permite analizar problemas y discutir resultados de su trabajo, lo cual contribuye a que se reduzcan la incertidumbre y los temores de los maestros ante la utilización

de nuevos métodos de enseñanza. Además, representa una oportunidad para autocriticar sus propias prácticas, discutir soluciones a nuevos problemas y encontrar y definir nuevos propósitos. Esta manera de mirar lo que hacen permite avanzar hacia soluciones específicas y experimentales y crear un marco de referencia para reestructurar los esfuerzos de las escuelas.

b) Los supervisores y coordinadores de microcentros locales focalizan su tarea de animación en una reunión mensual que también tiene carácter de seguimiento por cuanto en ella se revisa cómo se están aplicando las innovaciones. De esta manera dicha reunión se convierte en una fuente de valiosa información evaluativa. Desde 1985, este proceso de animación y seguimiento fue declarado como Proceso Oficial de Supervisión de la EN. Cada supervisor tiene la obligación de entregar un informe evaluativo sobre la marcha del microcentro y de cada uno de los maestros que lo integran.

c) Otro componente importante de la evaluación docente en la EN proviene de las investigaciones empíricas que se han realizado sobre el rendimiento que alcanzan los alumnos y los factores que lo determinan. Tanto los estudios de seguimiento como las opiniones de especialistas muestran que la EN ha mejorado las tasas de promoción, los puntajes de rendimiento y la autoestima, la satisfacción de los maestros y el apoyo de las autoridades locales y de la comunidad. De ello dan cuenta el informe que una misión de la UNESCO presentó tras una visita de observación y evaluación en 1985; la evaluación realizada por Rojas y Castillo del Instituto SER de investigación (1987), como la evaluación que en 1992 realizara el Banco Mundial en la que se califica la EN como una de las tres mejores experiencias en educación primaria.

d) Finalmente, las observaciones de algunos expertos que han visitado las escuelas y han podido observar directamente el desempeño de los maestros en sus aulas, así como las conversaciones sostenidas con ellos, también se convierten en información valiosa para la evaluación docente.

El programa de las 900 escuelas (P-900) de Chile

Al iniciarse el Gobierno de la Concertación Democrática en 1991, el Ministerio de Educación, que estableciera como ejes centrales de su política educativa «la mejora cualitativa de la educación, una mayor equidad para dar mayor oportunidad a los más necesitados y la integración de todos en la tarea educativa», comenzó el desarrollo de un programa que es conocido como P-900 por ejecutarse en las 900 escuelas más pobres del país. Las líneas de acción que se fijaron para el programa son las siguientes: el mejoramiento de la infraestructura; la entrega de textos escolares, bibliotecas de aula y materiales didácticos y la realización de talleres de perfeccionamiento para los docentes y de talleres de aprendizaje (TAPs) con participación de la comunidad.

El componente evaluativo sobre el desempeño docente puede ser percibido al

interior de los talleres de perfeccionamiento para docentes y que funcionan, al inicio, semanalmente en la escuela a cargo de un supervisor. El taller de capacitación para los docentes es realizado por cada supervisor en las mismas escuelas. La capacitación es grupal y permanente a lo largo del año, al principio con una reunión semanal, posteriormente cada quince días. Se intenta combinar la reflexión sobre la práctica docente con la introducción de nuevas metodologías. En general, además de trabajar unidades temáticas, estas reuniones sirven para fomentar el intercambio. Los docentes aprovechan el espacio para aliviar la tensión que les produce enfrentar las dificultades por las que atraviesan sus alumnos.

De esta manera se abre en las escuelas un espacio de reflexión pedagógica entre pares, sobre el tema de aprendizaje de los niños y sobre la propia práctica, para pensar en lo que se hace, evaluar sus resultados y buscar formas de hacerlo mejor (García-Huidobro, 1994). Ello trae como resultado que los docentes logren apropiarse de su práctica profesional, como un dominio en el cual pueden ser innovadores y creativos y sobre cuyos resultados tienen una responsabilidad profesional.

En esta línea de acción, destaca Filp (1994), se encarna uno de los principios orientadores del programa sobre la profesionalización del docente, que consiste en «apoyar al profesor en la toma de decisiones curriculares con base en sus propios criterios; proponerle transformaciones de sus prácticas más que tratar de imponérselas a modo de receta». En este sentido se considera que existe aquí una acción evaluativa del quehacer docente, que abarca tanto el ámbito pedagógico del aula como el de otras tareas administrativas de apoyo, sin que se observe en ello ningún tipo de resistencia por parte de los docentes.

Por otra parte, siguiendo el mismo estudio de Filp (1994), este componente evaluativo también abarca la acción de los mismos supervisores quienes, de una práctica que se caracterizaba por su aspecto fiscalizador, tras apropiarse en la capacitación que reciben de la propuesta pedagógica del programa, pasan a desarrollar un papel más técnico de supervisión. Esto es reconocido por la inmensa mayoría de supervisores, quienes afirman (97,5% respuestas de un cuestionario) que su participación les ha permitido enriquecer su práctica profesional.

En uno de los informes de evaluación externa y de acompañamiento que se hicieron al programa, Filp (1994) destaca la importancia de este componente de autoevaluación:

> La autoevaluación también jugó un papel importante en el permanente monitoreo de la calidad del trabajo realizado. Así, anualmente cada equipo de supervisión hace una evaluación del Trabajo Provincial, que es compartida en reuniones regionales en las que participan Jefes de los Departamentos Provinciales y Coordinadores del programa. En 1991 se elaboró una carpeta que le permite al cuerpo docente de cada escuela realizar una evaluación participativa de su labor educativa, actividad que se

desarrolla al finalizar el año.

Modelos que surgen desde la investigación

Existe una diversidad de enfoques evaluativos de la eficiencia docente, dependiendo de la finalidad que dicha evaluación persiga: se puede evaluar al profesor con fines administrativos, de promoción, retribución, ascenso o con fines didácticos, cuando lo que se busca es mejorar la calidad de su tarea pedagógica. Esta diversidad de modelos tiene que ver con las diferentes teorías de aprendizaje que están presentes en los sistemas y organizaciones existentes.

El objetivo de presentar los modelos de evaluación de la eficiencia docente que surgen de la investigación es el de facilitar un marco de referencia para comprender mejor la práctica de la evaluación de la acción docente en algunos países que han iniciado un proceso de reforma educativa.

Modelo centrado en el perfil del maestro

Consiste en evaluar el desempeño de un docente de acuerdo a su concordancia con los rasgos y las características—según un perfil previamente determinado—de lo que constituye un profesor ideal.

Estas características se pueden establecer de dos maneras. La primera consiste en establecer un perfil de las percepciones que tienen diferentes grupos (alumnos, padres, colegas) sobre lo que es un buen profesor. Puesto que todos hemos pasado por la escuela, cualquiera puede compartir su opinión sobre las características de los que ellos recuerdan como buenos profesores. A partir de las percepciones de estos distintos grupos (alumnos, padres y los mismos profesores) se construyen las características del «maestro ideal». Por supuesto que en este perfil van a aparecer rasgos referidos a ciertos aspectos de personalidad, como también otros referidos a aspectos técnico-pedagógicos. En estudios ya clásicos como los de Charters y Waples se recogen las características más altas: adaptabilidad, consideración, entusiasmo, buen juicio, honestidad y carisma. Los estudios más recientes realizados en esta línea como el de González Soler (1980) vienen a confirmar que «los profesores más eficaces son aquellos que se muestran más humanos en el amplio sentido del término, es decir, agradables afectuosos, justos, democráticos y más capaces de relacionarse con sus alumnos».

Otra manera de establecer el perfil del buen profesor es a partir de observaciones directas e indirectas, que permitan destacar rasgos importantes de los profesores que están relacionados con los logros de sus alumnos.

Una vez establecido el perfil, se elaboran cuestionarios que se pueden aplicar de diferentes manera, a saber: a manera de autoevaluación, mediante un evaluador externo que entrevista al profesor o mediante la consulta a los mismos alumnos.

Este modelo ha recibido críticas positivas y negativas. Un rasgo positivo es que la identificación del profesor se hace después de que la educación haya sido recibida; se deduce que los profesores que causaron mejor impresión han sido profesores eficientes y pueden ser identificados por los alumnos.

En cambio, un rasgo negativo sería que en los estudios de observación se sabe poco o no se sabe nada de los efectos a largo plazo de la tarea educativa, ya que este método de evaluación se basa en el proceso de enseñanza mientras ocurre. Además, los estudios de rasgos pocas veces distinguen entre las características humanas básicas que se dejan influir apenas por los procesos de educación de profesores y las características educables, como el conocimiento del contenido o la habilidad que tiene una persona para hacer preguntas. El aspecto más crítico del modelo es que establece el perfil de un profesor inexistente y cuyas características son prácticamente imposibles de inculcar a futuros maestros, ya que muchas de ellas se refieren a rasgos de carácter difícilmente enseñables mediante la capacitación.

También aparece como crítica negativa el que la gente no pueda llegar a un acuerdo referente a si una persona es un buen profesor y que haya poca relación entre las calificaciones y los logros de los alumnos. De acuerdo a estas consideraciones, este método no permite una evaluación objetiva y confiable sobre las características y competencias del profesor pues juzgar la efectividad de los profesores por los rasgos que demuestran, según Connelly (1990), es un asunto precario.

Por otra parte, si bien este enfoque ha sido criticado por construir sus indicadores sobre las percepciones de alumnos a los que no se les considera con condiciones para evaluar ciertos aspectos, otros investigadores como Ahumada (1992) mantienen que hay aspectos en los que la opinión del estudiante es indispensable y que éstos pueden proporcionar información válida y fiable.

Este modelo aporta instrumentos interesantes y ha contribuido a la aclaración de criterios, al mismo tiempo que permite establecer la relatividad del concepto de «buen profesor» por no poder conciliar los criterios o estándares que vienen desde perspectivas diferentes, del profesor y del alumno.

Modelo centrado en los resultados obtenidos

La principal característica de este modelo consiste en evaluar el desempeño docente mediante la comprobación de los aprendizajes o resultados alcanzados por sus alumnos. Surge de una corriente de pensamiento que es muy crítico sobre la escuela y lo que en ella se hace. Los representantes de este modelo sostienen que, para evaluar a los maestros, "el criterio que hay que usar no es el de poner la atención en lo que hace éste, sino mirar lo que acontece a los alumnos como consecuencia de lo que el profesor hace".

Al establecer este criterio como fuente esencial de información para la evaluación del docente se corre el riesgo de descuidar la calidad de la educación, pues, como

muy bien señala Ahumada (1992), "al saber el docente que se le está evaluando por los logros estudiantiles, es un hecho que enfatizará en ellos el aprendizaje de conductas reproductivas de más bajo nivel, en desmedro de conductas superiores que, de por sí, son más difíciles de alcanzar". Por otra parte, hacer un planteamiento de tal naturaleza es poco sólido porque no se puede hacer considerar al profesor como responsable absoluto del éxito o aprendizaje de los alumnos. El estudio de Cardemil (1991) es elocuente al señalar que los resultados que obtienen los alumnos son efectos de múltiples factores, uno de los cuales sería el docente. Se diría que este modelo desmerece el trabajo del maestro, al no reconocer la complejidad de la tarea de enseñar y limita el concepto de aprendizaje al centrarlo en la transmisión y reproducción de información.

Modelo centrado en el comportamiento del docente en el aula

En este modelo se pone énfasis en la enseñanza como un proceso y su correlación con variables de productos tales como los logros de los alumnos. El modelo propone que la evaluación de la eficacia docente se haga identificando aquellos comportamientos del profesor que se consideran relacionados con los logros de los alumnos. Dichos comportamientos se relacionan, fundamentalmente, con la capacidad del docente para crear un ambiente favorable para el aprendizaje en el aula. Este modelo ha predominado desde la década de los años sesenta, empleando pautas de observación, tablas de interacción o diferentes escalas de medida del comportamiento docente.

La crítica que recibe esta forma de evaluación se refiere más bien a la persona que realiza la evaluación. Se objeta que los registros obedecen a la concepción que los observadores sostienen sobre lo que es una enseñanza efectiva y que se demuestra por los estándares que sustentan para cada hecho observado. La subjetividad del observador entra fácilmente en juego y posibilita que éste gratifique o perjudique a los observados por razones ajenas a la efectividad docente, sino más bien por su simpatía o antipatía hacia ellos.

Por otra parte, Ahumada (1992) mantiene que el conocimiento anticipado de los criterios, evidencias y estándares empleados en las pautas puede llevar a los maestros a modificar su forma de actuar en el aula, desaprovechando en algunos casos cualidades personales favorables de interacción y disminuyendo, por ende, su efectividad en la docencia. La aplicación de este modelo contribuye al aporte de técnicas e instrumentos de observación siempre que se haga a partir de la observación empírica y de los propios enfoques educativos que cada país desarrolle.

Modelo de la práctica reflexiva

Consiste en una instancia de reflexión supervisada. Se trata de una evaluación para la mejora del personal académico y no de control para motivos de despidos o promoción. En ese sentido aporta informaciones que permiten tomar decisiones para el mejoramiento de la calidad de la enseñanza. El modelo se fundamenta en

una concepción de la enseñanza como «una secuencia de episodios de encontrar problemas y resolver problemas, en la cual las capacidades de los profesores crecen continuamente mientras enfrentan, definen y resuelven problemas prácticos» a la que Schön (1987) llama reflexión en la acción y que requiere de una reflexión sobre la acción o evaluación después del hecho para ver los éxitos, los fracasos y las cosas que se podrían haber hecho de otra manera.

La ejecución de este modelo contempla tres fases o etapas, a saber:

- Una sesión de observación y registro anecdótico de la clase.
- Una conversación reflexiva con la persona que observa para comentar lo observado y en la que se hacen preguntas encaminadas a descubrir la significatividad y la coherencia de la práctica observada.
- Una conversación de seguimiento en la que se retoman los temas conversados y las acciones acordadas en la segunda fase. Si es necesario y conveniente, en esta fase se puede hacer una nueva observación con registro.

La aplicación de este modelo requiere de la existencia de un sistema de supervisión, con personas y tiempos destinados a ello. Sin embargo, el modelo puede ser adaptado para que la observación sea hecha por otras personas como, por ejemplo, colegas del mismo establecimiento o algún docente directivo.

Por ser un modelo que además de la flexibilidad, resalta lo participativo y educativo—puesto que el uso de la información es para fortalecer o transformar las prácticas—se presta para ser usado también con las personas en proceso de formación para ser maestros.

Evaluación del docente y asociación gremial

El trabajo docente en América Latina, desde el punto de vista del gremio, se caracteriza por dos fuertes tendencias: la que enfatiza en el trabajo docente el proceso de «proletarización» y la que acentúa su carácter «profesional-técnico».

A su vez, las organizaciones gremiales del magisterio en Latinoamérica están marcadas por una politización que se justifica debido a situaciones propias de la región. Éstas se refieren, por un lado, a: una presencia fuerte del Estado en el mantenimiento y la gestión educativa que lo convierte en patrón y legislador y, por otro, a la incidencia de los partidos políticos que se convierten en intermediarios ante los poderes públicos, dada la relativa debilidad de los movimientos gremiales. Esto se ve acentuado por un fuerte sentido reivindicativo que surge como consecuencia de las condiciones laborales en las que el magisterio vive y desarrolla su acción. De acuerdo al estudio de Núñez (1990), tales características hacen que se dé una tendencia en los Estados a excluir a las organizaciones gremiales de la toma de decisiones para las políticas educativas.

La preocupación del sindicalismo docente se ha centrado más en los temas

reivindicativos, institucionales y administrativos que en aquéllos que tienen que ver con la orientación del curriculum o lo propiamente pedagógico.

Las organizaciones gremiales no se han servido de los aportes de la investigación por desconfianza a un supuesto carácter tecnocrático. Las reformas educativas no han abierto espacios para conciliar la participación creativa de las organizaciones. Es más, en el mismo estudio de Nuñez se señala que en muchos casos la reacción ante las reformas ha sido de oposición. Son pocas las experiencias en que las organizaciones no han adoptado una actitud reactiva, de regateo o rechazo o carecido de propuestas alternativas.

Un caso es la Asociación General de Profesores de Chile, que entre 1923 y 1927 hizo la crítica de la educación y levantó su propio proyecto orgánico y pedagógico que llegó a traducirse en una efímera, aunque significativa, reforma general de la educación en 1928.

Otro, es el caso colombiano de la Federación Colombiana de Educadores (FECODE) que aparece como uno de los aportes más interesantes y novedosos desde el movimiento gremial. Mejía (1987) explica cómo el Movimiento Pedagógico (que surge al interior de las asociaciones gremiales) logra, además de una transformación interna del sindicato, "dar comienzo a un proceso a partir del cual el maestro se critique, que no se tienda un manto protector desde la organización, que se le exija una seriedad profesional que le permita una identidad en su profesión, (...) que le deje pensar en lo que ha sido su devenir histórico y reconstruir la historia de su práctica como principio de identidad con sus acciones."

El Movimiento Pedagógico de Colombia propone a los maestros dejar de ser reproductores para convertirse en constructores de la sociedad. Esta tarea implica asumir una acción de evaluación que no encuentra oposición ya que viene desde el interior de su quehacer cotidiano y se refiere, no a aspectos administrativos ni organizativos de su función, sino a algo más medular como es el ser artífice de la construcción de conocimientos para la sociedad. En el Movimiento Pedagógico hay una conciencia clara de que esto supone un aprendizaje pues, al devolver la voz a los maestros, éstos tienen que iniciarse en el oficio aprendiendo a "balbucear sus primeras palabras sobre su quehacer, rompiendo los miedos que aquello implica" (Mejía, 1987). Y uno de esos miedos se refiere a que el maestro se atreva a hablar a su nivel evitando ser suplantado por los "expertos" de la educación. Pero sobre todo el aprendizaje se refiere a aceptar la evaluación como un componente necesario para su función y que se reconoce en los lineamientos del Movimiento Pedagógico, especialmente en lo que dice: "Que el maestro piense su práctica, la discuta, e inicie un proceso permanente de crítica y autocrítica".

Lo que los movimientos pedagógicos aportan es una práctica reflexiva que permita a los docentes pasar de un "rol meramente técnico" a un "rol profesional" pues lo que ellos promueven es un proceso de reflexión que, más que atiborrar de discursos, lleva a los maestros a reflexionar sobre su práctica, a analizar cómo se enseña, a

escudriñar cómo se fijan las normas y las interacciones y a descubrir ese "para qué" se enseña.

Otros movimientos pedagógicos donde también ocurren procesos semejantes al caso colombiano incluyen: el grupo Simiente en el Brasil, las Escuelas Antonio Encinas del Perú y los grupos Espacio y Freinet en Chile.

PRÁCTICA DE LA EVALUACIÓN DOCENTE

Para presentar algunos casos concretos de evaluación de la acción docente se han elegido aquellos países sobre los que existía información en REDUC. Posiblemente existen otras experiencias sobre las que no hay documentación escrita en esta red. No obstante, lo que muestra cada uno de los casos recopilados en este trabajo es un abanico amplio de tendencias, realidades y aspiraciones que bien pueden ayudar a estudiar y sugerir líneas en las que hay que seguir profundizando.

Argentina (Provincia de San Juan)

Desde 1976 existe la Ley 2492, Estatuto del Docente, que habla de "la calificación de los docentes" (Capítulo XV, artículos 82 al 85) refiriéndose a la actuación profesional de los docentes, que se irá registrando en un "legajo personal de actuación" y que el docente tiene derecho a conocer e incluso a poseer un duplicado de toda la información allí registrada.

La calificación, que debe ser anual, está a cargo de una Junta Clasificadora y contempla los siguientes aspectos: cultura general y profesional, aptitudes docentes y disciplinarias y asistencia y puntualidad. Existe una ficha tipo para emitir el concepto profesional. La ley prevé también mecanismos de apelación en caso de que el profesional no esté de acuerdo con el concepto que se emite sobre él. En reiteradas oportunidades se ha querido modificar este Estatuto pero lo único que se ha logrado es introducir en el mismo uno que otro cambio.

A pesar de la reglamentación, la práctica de la evaluación parece ir por otro lado y, según el sentir de los maestros, no incide en la mejora del desempeño docente. Además, genera muchos conflictos que se resuelven otorgando a todo el mundo el puntaje máximo, convirtiendo la evaluación en un rito por confabulación mutua que no mejora nada, o entrando a desarrollar un largo proceso de apelación cuando hay una calificación pero no hay elementos para fundamentarla.

Según la Ley, el "equipo directivo" debe asumir la tarea de evaluación en cada establecimiento, pero en la práctica esto es asumido sólo por el Director quien luego la hace llegar a la Junta Clasificadora con todos los riesgos que esto implica[2].

Hay un sentir bastante general en los maestros acerca de que la Planilla de Conceptos que se utiliza para la calificación anual "no sirve para nada». Es más, cuando en reuniones de Supervisores se aborda la necesidad de un seguimiento o

monitoreo para tener más información de cómo se desarrollan las clases, éstos encuentran como respuesta: "los directores no tienen tiempo".

Dentro de la Universidad Nacional de San Juan, de la que dependen varios colegios, se está trabajando en un cambio de la normativa sobre calificación. Y algo que los especialistas y grupos que están a favor de la innovación vienen recomendando es que en los establecimientos se formen equipos docentes de cursos paralelos y se les capacite para realizar observaciones mutuas de sus clases.

El Movimiento Gremial de Docentes (UDAP:Unión de Docentes Agremiados Provinciales; UDA: Unión de Docentes Argentinos; y AMET: Asociación del Magisterio de Educación Técnica), amplio y variado, parece no tener una postura oficial al respecto. Este es un tema que cae dentro de lo profesional y la preocupación mayor del gremio, en la coyuntura actual, es lo salarial. Actualmente, ante la posibilidad de que el Gobierno quiera evaluar a los docentes con base en la «productividad» hay una oposición fuerte del gremio, el cual defiende que no se pueden utilizar para la evaluación del docente indicadores como «cuántos pasan de curso, asistencia,etc».

Colombia

La Ley General de Educación (Ley 115, del 8 de febrero de 1994), aprobada por el Congreso de la República dentro del marco de la nueva Carta Política, ordena que debe construirse un Sistema Nacional de Evaluación de la Educación con el fin de velar por la calidad, el cumplimiento de los fines de la Educación y la mejor formación de los educandos. Señala que el objetivo de dicho Sistema será diseñar y aplicar criterios para evaluar la calidad de la enseñanza, el desempeño de los docentes y directivos, los logros de los alumnos, la eficacia de los métodos, textos, materiales y la organización administrativa. Establece dos mecanismos de evaluación: cada año se llevará a cabo una evaluación del establecimiento educativo; y cada seis años, los educadores deben rendir un examen de idoneidad académica en el área de su especialidad docente y de actualizaciones pedagógica y profesional (arts. 80, 81 y 84).

Es importante destacar que tanto a los docentes como a los directivos, en el caso de que no alcancen el puntaje requerido, se les otorga una nueva oportunidad. En el caso de los docentes, éstos pueden volver a presentarse a examen en el tiempo máximo de un año. En el caso de los directivos, también tienen un año para que presenten un proyecto que tienda a solucionar los problemas encontrados (arts. 81 y 82).

La Ley dispone que este Sistema opere en coordinación con el Servicio Nacional de Pruebas del Instituto Colombiano para el Fomento de la Educación Superior (ICFES). El artículo 105 de la Ley establece también las normas de incorporación de los docentes al servicio estatal y se prevé la situación de los que están actualmente vinculados, respetando su estabilidad laboral. Para la puesta en marcha

de lo establecido por la Ley, tanto el ICFES como otros organismos han comenzado a hacer propuestas y a elaborar proyectos dentro del marco legal establecido.

Según el Proyecto de Evaluación de Docentes en Ejercicio preparado por la Universidad Nacional de Colombia (1996), la evaluación ha de entenderse como un proceso y el examen como un momento dentro del mismo que permite establecer un diagnóstico para la definición de estrategias de mejoramiento de la calidad. Este proyecto señala que una evaluación de la labor del maestro debe contemplar su acción educativa como una totalidad y no limitarse a un aspecto parcial de su trabajo. Es decir, la evaluación del trabajo de los docentes debe pensarse como un fundamento para la elaboración de planes y programas de mejora educativa y no como una estrategia de clasificación que pueda conducir a medidas de exclusión o de castigo. Por ello, la evaluación debe ubicarse en el complejo territorio de la práctica pedagógica entendida como proceso que integra información proveniente de distintas fuentes y como "competencia pedagógica" que tiene:

- una **dimensión comunicativa:** se refiere al modo como circula el contenido conceptual e informativo de las interacciones pedagógicas y a los lenguajes que la hacen posible;
- una **dimensión ética:** se refiere a los paradigmas de comportamiento que facilitan el desarrollo de un esquema de valores propios de una sociedad democrática y correspondiente a una cultura académica;
- una **dimensión estética:** tiene que ver con la capacidad de asombrarse y el modo de acompañar a los estudiantes en su proceso de descubrimiento; no puede estudiar sino examinando las estrategias pedagógicas mismas; y
- una **dimensión psicosocial:** se refiere a la capacidad del maestro para aportar al niño lo que necesita sin que deba renunciar a lo que le entusiasma.

Si la reflexión del maestro sobre su propia práctica pedagógica es fundamental para cualificar dicha práctica, no es menos cierto que esta reflexión es mucho más rica si se la compara con la de otros docentes. De ahí que sea pertinente que los docentes se examinen y evalúen unos a otros. El juicio de los estudiantes constituye también una importante fuente de conocimiento de la relación pedagógica.

Los criterios, pues, que conforman un marco para un sistema de evaluación y para los instrumentos que se usen se referirán a:

a) La articulación del mundo escolar y el extraescolar. Lo que significa que debe darse un proceso de adaptación de los conocimientos decantados por la ciencia y la cultura para volverlos significativos y comprensibles en el contexto escolar y en la vida cotidiana de la comunidad. Además, implica selección, jerarquización, reconstrucción y reorganización de tales conocimientos.

b) Dominio de lenguajes escolares. Es decir, que la labor del maestro exige una capacidad de interpretación de textos escritos y de códigos diversos, capacidad de

interpretar el lenguaje y los dibujos de los niños y tener una interacción con ellos; y que promueva una ética de la comunicación, lo cual implica un compromiso con la palabra.

c) Capacidad de innovación. O sea que, por un lado, el maestro manifieste capacidad para adaptar textos y programas a los contextos y a las situaciones particulares del aula y, por otro, que se den las condiciones para estimular innovaciones por parte de los maestros.

d) Compromiso social. La escuela es una instancia ineludible de socialización, lo que confiere al maestro una responsabilidad social. Se hace necesario que el docente comprenda las implicaciones de su práctica, así como las limitaciones que la misma sociedad, la comunidad y el estado le imponen.

El Sistema de Evaluación busca promover una apreciación, desde distintos ángulos o miradas, de la competencia del maestro. Siguiendo los criterios antes señalados, se estructura con los siguientes componentes:

1) Autoevaluación. Su propósito es promover la reflexión del maestro sobre su propia práctica, sobre las condiciones que la potencian o limitan y sobre su incidencia en la sociedad y la comunidad, así como indagar sobre ciertos aspectos como la formación y actualización; la calidad de su labor de enseñanza, de proyectos institucionales, textos y programas; la calidad de la interacción con los estudiantes; las dificultades (de infraestructura, de la propia formación, económicas o culturales) para la realización de su labor, y las relaciones con otros compañeros, los directivos y la comunidad.

2) Prueba de competencias pedagógicas y gramáticas de la escuela. Permite reconocer la capacidad para adaptar conocimientos, métodos y estrategias a las circunstancias en que trabaja el maestro y el dominio de los fundamentos teóricos de su especialidad.

Las Gramáticas Generales de la Escuela se refieren a la competencia en interpretación de textos escritos y otros códigos (matemáticos, gráficas, esquemas, etc.), en la organización del espacio, en la observación sistemática y en la realización de operaciones diversas de medición, que son capacidades que todo maestro debe poseer, independientemente de la especialidad y curso que enseñe. Por su parte, las Gramáticas Específicas de la Escuela se refieren a la comprensión y puesta en práctica de los principios básicos, métodos y estrategias de cada disciplina.

3) Pregunta abierta. A través de la redacción de un ensayo corto, sobre un tema relacionado con su práctica, se busca que el maestro ponga en juego tanto su conocimiento teórico y su capacidad de argumentación, como su propia práctica. El trabajo será evaluado por los propios maestros de acuerdo a criterios tales como la coherencia del texto, la fuerza de argumentación, la riqueza de la experiencia y la

capacidad de expresión y redacción.

4) Evaluación por parte de los colegas. Persigue indagar la capacidad de los docentes para fomentar y adelantar el trabajo colectivo dentro de su institución. Se lleva a cabo en forma colegiada.

5) Evaluación por parte de los estudiantes. Busca precisar la forma en que el docente interactúa y se comunica con sus alumnos, así como el compromiso del docente con su trabajo.

6) Encuesta. Explora los recursos disponibles y su utilización, las relaciones escuela-comunidad y los énfasis temáticos y metodológicos para contextualizar el trabajo educativo. Ello permite la conformación de un mapa pedagógico para las distintas regiones y contextos.

Para la puesta en marcha de este examen de idoneidad se han elaborado nueve tipos de examen: un tipo único para el nivel Básica Primaria y ocho tipos diferentes para Básica Secundaria con énfasis en la especialidad.

La estructura del examen presenta tres componentes: a) concepciones de lo educativo (aproximadamente 50 preguntas); b) gramáticas de la escuela generales, iguales para todos los tipos de exámenes (aproximadamente 30 preguntas); y c) gramáticas de la escuela específicas en el área de especialidad docente, únicas para cada uno de los nueve tipos de exámenes de este Nivel de Educación Básica (aproximadamente 30 preguntas).

El puntaje máximo de la prueba será el promedio obtenido por los docentes que se ubiquen en el tramo del 20 por ciento de las calificaciones más altas. El puntaje mínimo será el 20 por ciento del puntaje máximo.

Mientras se reglamenta el artículo 105 de la Ley General de Educación relativo al Sistema de Selección de Docentes para su ingreso al sector educativo, se está actuando de acuerdo a los procedimientos establecidos anteriormente y a una resolución de 1989 que establece las bases, las formas y los procedimientos de los concursos. Éstos son convocados a niveles distrital y departamental, pero el Ministerio de Educación cuenta con un Grupo de Concursos y Evaluación Docente para prestar asistencia técnica y asesoría administrativa y pedagógica al respecto.

La participación en los concursos se hace mediante:

- una prueba que evalúa aspectos de cultura general, aptitud pedagógica y administrativa y conocimientos pedagógicos y académicos, por área. Esta es una fase de carácter eliminatorio que tiene un valor del 60 por ciento de la calificación final;
- una entrevista, igualmente eliminatoria, con valor de un 20 por ciento;

- entrega de antecedentes, a los que se da la siguiente valoración: 5 por ciento si es oriundo del municipio; 5 por ciento por experiencia de cinco años o más; 10 por ciento si su experiencia es en el sector rural.

Costa Rica

Dentro del sistema educativo de este país, el área de evaluación del desempeño docente es reconocida como prioritaria en la administración por cuanto sirve para verificar el aporte de los docentes en la obtención de los objetivos educativos que persigue el sistema. Prueba de ello son los numerosos estudios realizados al respecto, entre los que cabe destacar:

- El Perfil Psicoprofesional propuesto por el Departamento de Personal del Ministerio de Educación Pública (MEP) realizado en 1969. Considera los aspectos personales y profesionales del docente y determina los criterios para evaluarlos.
- La investigación acerca de la Validez del Sistema de Calificación Profesional de Docentes que realizó el Centro Multinacional de Investigación Educativa (CEMIE) en colaboración con la Unidad de Investigaciones Pedagógicas del Ministerio de Educación Pública en 1974. Permite establecer y describir las características deseables de los docentes en Costa Rica.
- El Departamento de Investigación Educativa del Ministerio de Educación Pública inició en 1983 un plan dirigido al control de la calidad; dicho plan también comprendía el desarrollo de estudios orientados a la elaboración de perfiles, llegándose a establecer un ideal de maestro urbano, maestro unidocente y maestro rural; posteriormente se realizaron una serie de estudios relacionados con la determinación de las características personales y profesionales que deben poseer los profesores de distintas asignaturas científicas, ciencias sociales y lenguas.

Las primeras informaciones que se tienen acerca de la evaluación del educador costarricense aparecen en el Capítulo V del Código de Educación, promulgado en 1944. Este capítulo hace referencia a una hoja de servicios para calificar la labor docente durante el año lectivo y que serviría para efectos de ascenso, pensión y aquellos otros que consigna el mismo código.

Según los artículos 109 a 113, la hoja de servicios debía ser cumplimentada por el mismo director del establecimiento donde trabaja el docente una vez al año, al final de cada curso lectivo. Contenía información referente al tiempo de servicios durante el año, la puntualidad en la asistencia, la calidad de servicios, estímulos y censuras.

Además, se estableció una Junta Calificadora del Personal Docente presidida por el Jefe Técnico de Educación Primaria e integrada por vocales propietarios. Tenía entre sus atribuciones asignar categorías a los maestros, acordar los ascensos y formular el escalafón general del maestro.

La Ley Fundamental de Educación de 1957 recoge los fines de la educación costarricense, la forma como se organiza el sistema educativo y los ciclos que lo conforman. Se señala en la ley que el Ministerio de Educación será el encargado de elaborar los proyectos de las leyes y los reglamentos derivados de las mismas. También se propone una Ley de Personal Docente inspirada en la concepción democrática de la educación pública que deberá establecer, entre otras cosas, la evaluación de la labor del personal[3].

En 1969, siguiendo recomendaciones de la comisión que elaboró la Ley Orgánica del Ministerio de Educación Pública, el Departamento de Personal de ese Ministerio confeccionó el Manual para la Evaluación y Calificación de Servicios Profesionales de la Carrera Docente, el cual propone la aplicación de un sistema de evaluación y calificación del personal incluido en la carrera docente.

Se establece, pues, un sistema que consta de dos elementos fundamentales:

- un formulario modelo para evaluar a cada uno de los funcionarios incluidos en la carrera docente; y
- un Manual para la Evaluación y Calificación de los servicios Profesionales de la Carrera Docente.

Sin embargo, ni los que deben ser evaluados, ni los que realizan la evaluación están contentos con lo que se hace ni con los instrumentos utilizados. Los resultados de la investigación de Fallas, Herrera, Páez y Zamora (1993) sobre el Sistema de Calificación del Servicio Civil para Docentes muestran que « se observa un generalizado error de lenidad descrito como la tendencia de ciertos observadores a sobreestimar los sujetos que juzgan. La frecuencia de puntaje excelente es claramente mayor que la esperada en una distribución normal. El director no se atreve a descalificar sino en casos muy obvios» y, además, que «se infiere la existencia del efecto de halo, en el que el evaluador se forma una impresión global del evaluado y, a partir de ella, asigna puntos a los diferentes aspectos. Es decir, no se evalúa diferencial e independientemente cada aspecto».

En 1979 la Comisión Nacional de Planificación y Programación denunció que el sistema engendraba numerosos problemas y pidió que se revisaran los procedimientos. En 1983 se modificó nuevamente la Hoja de Evaluación y Calificación de Servicios. Tanto de parte de los especialistas en educación como de los docentes asociados en gremio, se escucharon críticas con respecto a que los cambios habían sido pocos y más de forma que de fondo: en 1983 se cambió el tamaño de la Hoja por uno menor y se agregaron casillas en las que el evaluador anotaba la calificación que correspondía por área; en 1985 se agregaron algunas instrucciones para llenar el formulario y se volvió a detallar cada aspecto a evaluar. Sin embargo, las organizaciones magisteriales se mostraban interesadas en buscar un consenso con el Ministerio de Educación para actualizar la Ley de Carrera Docente e impulsaron un proyecto de reforma en lo referente a la Evaluación y Calificación

de Servicios, pero las modificaciones que proponían eran también de forma.

En el trabajo de Sánchez (1992) se muestran los resultados de una encuesta formulada al personal docente y que forma parte de la propuesta de un modelo de evaluación y calificación de servicios para el personal docente. Según estos datos, parece existir una gran aceptación (83,3% de los docentes), respecto al uso de los conceptos de evaluación: excelente, muy bueno, bueno, insuficiente e inaceptable. De igual forma, un 64% de los encuestados están de acuerdo en practicar la autoevaluación. Hay consenso entre los docentes en mantener los factores comprendidos en el instrumento. Para algunos servidores el instrumento de calificación de servicios «no es malo», lo que falla es su utilización. A pesar de esto, el 84% manifiesta que el instrumento se les ha aplicado con objetividad la mayor parte de las veces.

Sin embargo, la mayoría de los consultados desconocen el objetivo de la evaluación y la calificación de servicios según señala la Ley de Carrera Docente y el Reglamento de Carrera Docente, a saber, «la concesión de estímulos y beneficios tales como traslados, ascensos, licencias para estudios, aumentos anuales, becas y facilidades».

Chile

Chile está en la mira de todos por sus logros en materia de desarrollo y el impacto social de programas educativos de reforma. Se esperaría que en materia de evaluación del desempeño docente también tuviera experiencias exitosas que mostrar. Al igual que en otros casos, tal vez sea importante considerar los elementos que están en juego en la búsqueda.

Un primer elemento se refiere a la evolución que sobre la regulación legal del trabajo docente ha ido ocurriendo en los últimos veinte años. Según Núñez (1996), el docente ha pasado de un régimen tradicional de funcionario estatal a otro de descentralización de la gestión de personal, en el que se intenta una búsqueda de equilibrio entre las diversas regulaciones nacionales encaminadas a garantizar la equidad y el respeto de los derechos legítimos de los profesionales de la educación. Durante esta transición el tema de evaluación del desempeño docente puede decirse que ha sido tangencial pues antes que normar sobre esto, tanto a las autoridades estatales como al mismo gremio de docentes, les interesaba llegar a establecer condiciones más dignas de trabajo.

En 1980 los trabajadores de las escuelas y liceos públicos dejaron de ser funcionarios del Estado y se les traspasó a las Municipalidades, con contrato como trabajadores según la ley laboral común. Esto trajo como consecuencia la pérdida de derechos, como el de la estabilidad laboral, por parte de los docentes, comenzando a establecerse altas diferencias de remuneraciones.

Todo ello dió lugar a una fuerte movilización del gremio para buscar mecanismos

que le permitiera mejorar sus condiciones laborales. En 1990, con el restablecimiento de la democracia, el movimiento sindical docente se fortaleció en demanda de un Estatuto propio que logró sus frutos en 1991 cuando se dictó el Estatuto de los Profesionales de la Educación, el cual:

- integra bajo una misma matriz legislativa a los docentes de todos los niveles y modalidades del sistema público y del sector particular;
- mejora sustantivamente las condiciones de estabilidad laboral de los docentes al permitir la titularidad del cargo, automáticamente, a todo profesional que en 1991 estaba con contrato indefinido;
- permite los traslados y las permutas en el sector municipal sin perder los reconocimientos a la antigüedad y al perfeccionamiento adquiridos como sucedía antes;
- regula las jornadas de trabajo, horarios máximos y distribución interna de éstos. Legaliza las vacaciones de verano y establece que los docentes con más de 30 años de servicio tienen derecho a disminuir su carga docente sin disminución de remuneraciones;
- no es discriminatorio con el docente que trabaja en el sector particular sino que, por el contrario, abre perspectivas de equiparación como nunca antes se dieron.

El debate en torno a la puesta en práctica del Estatuto ha estado centrado en estos aspectos de mejoramiento de las condiciones laborales en desmedro de aquellos otros aspectos que atañen más a la calidad, como es el caso de la evaluación del desempeño docente sobre el que se ha sentido una permanente resistencia a su aplicación. Al respecto el Reglamento de la ley (art.59, párrafo XI) establece que:

> Los profesionales de la educación son personalmente responsables de su desempeño en la función correspondiente. En tal virtud deberán someterse a los procesos de evaluación de su labor y serán informados de los resultados de dichas evaluaciones. Los profesionales de la educación gozarán del derecho a recurrir contra una apreciación o evaluación directa de su desempeño, si la estiman infundada. La calificación tendrá por objeto evaluar ese desempeño funcionario del profesional de la educación para los efectos que señala el presente Reglamento.

El Comité Técnico Asesor del Diálogo Nacional sobre la Modernización de la Educación (1994) enfatiza, en el documento presentado al Gobierno, la necesidad de continuar con los esfuerzos de mejoramiento de las condiciones laborales a fin de profesionalizar más al docente. Pero también invita, dentro de este contexto de profesionalización, a adoptar esquemas que permitan introducir "adecuados incentivos de retribución al trabajo realizado, reconocimiento del esfuerzo personal y premio al buen desempeño y sus resultados". Específicamente menciona el esquema puesto en marcha en otros países como es el «pago por méritos».

Durante todo este tiempo la postura de los docentes, respaldada por el Colegio de

Profesores, ha sido de resistencia a todo intento de evaluación del desempeño docente. Ellos opinan que la propuesta evaluativa sugerida por el Estatuto Docente es inconveniente porque se basa en una calificación punitiva. Se dice que los profesores son evaluados por factores ajenos a su calidad de profesores. Tales motivos llevaron a rechazar la propuesta del Estatuto y a buscar la negociación de otra propuesta.

Actualmente se busca la concreción de una propuesta evaluativa en una comisión bipartita Ministerio de Educación-Colegio de Profesores. Desde el Colegio se piensa que la propuesta que se estudia considera importantes innovaciones como por ejemplo la evaluación separada por estamento (Director y docentes administrativos; integrantes de las Unidades Técnico Pedagógicas (UTPs); docentes propiamente tales).

Se busca evaluar la actitud y condición de trabajo frente al entorno, a la vez que premiar el liderazgo y la actividad. Por ejemplo, la evaluación del par evalúa la actividad en la comunidad escolar. Es importante que el director no evalúe solo sino que recopile la información y consulte antes de realizar la evaluación. Para este caso, se propone también llevar un registro de vida del profesor con conocimiento del mismo.

La finalidad de esta evaluación es clasificar al profesorado en 5 conceptos o categorías, además de tener una lista de méritos (importante para cuando el profesor se presenta a concurso público, solicita una beca o como incentivo monetario) y una lista de demérito (fase del proceso de mejoramiento. Si por dos años consecutivos está en esta lista, el profesor debe salir del sistema público). Según el gremio, el proyecto ya debería estar en marcha.

La actitud de resistencia expresada se manifiesta con claridad en los estudios realizados sobre la percepción que tienen los docentes de la supervisión pedagógica que, según el Manual de Supervisión Técnico Pedagógica publicado por el Ministerio de Educación en 1990, realiza una función evaluadora. En el estudio de Rubilar y Cuevas (1996) se da cuenta que los docentes no están conformes con la supervisión pedagógica ejercida por los supervisores externos o los de las Direcciones Provinciales de Educación. Consideran que, además de las escasas visitas, no reciben de ellos ni orientaciones ni sugerencias pertinentes, ni contribuyen a incentivar el trabajo docente. En cambio, valoran las acciones de supervisión de los Jefes Técnicos o supervisores internos de sus establecimientos con quienes pueden tener una permanente relación y con gusto aceptan sus sugerencias.

Por otra parte, las autoridades educativas de gobierno, con el afán de llevar adelante la política educativa propuesta con el restablecimiento de la democracia—cuyos principales objetivos eran mejorar la calidad de la educación, lograr equidad en su distribución y perfeccionar la participación de los diversos agentes sociales en la tarea educativa —ha ido impulsando diversos programas o acciones que han podido

integrar la innovación y el fortalecimiento de la profesionalización docente en cuanto en dichos programas está el aspecto de evaluación del desempeño como un componente necesario para el desarrollo del mismo.

Tal es el caso de los Proyectos de Mejoramiento Educativo (PMEs) elaborados por los mismos docentes del establecimiento a partir de lo que ellos consideran importante y necesario hacer para renovar y mejorar el aprendizaje de sus alumnos. Todo ello les obliga a mirar su propia práctica, enjuiciarla y buscar alternativas que signifiquen un mayor éxito en sus alumnos y una satisfacción de ellos como profesionales al estar ejecutando proyectos que ellos mismos han elaborado. Entre 1992 y 1995 se han ejecutado 3.111 PMEs, lo que significa que más de 75.000 profesores han participado en su elaboración. De acuerdo al Directorio de Proyectos de Mejoramiento Educativo del Ministerio de Educación (1996), en 1995 había 1.680 escuelas básicas con PMEs en ejecución.

Enfoques muy semejantes se dan en otros programas en los que está presente el componente de evaluación como un aspecto que ayuda a la transformación de las prácticas y a la mejora de la calidad. Sin embargo, esto no tiene incidencia en el puntaje, reconocimiento o incentivo profesionales.

No existe, sin embargo, normativa respecto a la evaluación del docente. Sí la hay sobre la evaluación de los establecimientos y a partir de ella se deduce una evaluación de los docentes. Las autoridades educativas están buscando el mecanismo para llevar a cabo una evaluación también de carácter individual. A ello apunta la creación del Sistema Nacional de Evaluación de los Establecimientos (SNED) que asigna incentivos a la calidad del trabajo colectivo. Se intenta tener en cuenta, no sólo los resultados académicos, sino también otros factores como la iniciativa o innovación, los grados de participación y vinculación con el medio, y cómo cada escuela contribuye a la igualdad de oportunidades. Además, se está buscando el desarrollo de políticas de incentivo a los establecimientos y a su personal.

En cada establecimiento el incentivo es de igual valor para todos los profesionales de la educación que en él trabajan, en proporción al número de horas semanales de su contrato. En consecuencia, se trata de un incentivo al desempeño colectivo. Sin embargo, la ley reserva un 10% del financiamiento del incentivo entregado a cada escuela para que su personal pueda decidir autónomamente un nuevo reparto igualitario de esa reserva o decidir asignarlo a algún o algunos profesionales que el colectivo encuentre particularmente destacados. En esta disposición está contenido el principio de un incentivo individual.

Por otra parte Núñez (1996), en un documento interno del Ministerio de Educación, señala los propósitos que la Comisión conjunta Colegio de Profesores-Ministerio de Educación tiene para el estudio de la reforma del párrafo del Reglamento del Estatuto Docente que se refiere a las calificaciones del personal profesional. Tales propósitos son: hacerlo menos punitivo, más estimulante del

desarrollo profesional, reducir toda posibilidad de arbitrariedades, enriquecer los instrumentos, ampliar el número de los participantes en los procesos de evaluación, hacer que las buenas calificaciones sean el mejor recurso de protección a la estabilidad y las bajas calificaciones reiteradas permitan tomar decisiones que protejan y promuevan el interés de una educación de calidad.

No se conocen—y eso llama la atención—registros escritos de experiencias realizadas sobre evaluación individual del desempeño docente. Pero se sabe que en algunos establecimientos, especialmente en el sector privado, se han realizado intentos con muy escaso éxito, lo cual ha llevado a los mismos docentes y a la organización gremial a seguir manteniendo una postura reticente a que se legisle sobre esta materia.

El Salvador

El proceso de transformación educativa que vive este país es altamente dinámico a pesar de la compleja situación socioeconómica a que se enfrenta fruto de la guerra y la posguerra, esforzándose por consolidar una firme paz.

Tanto los estudios de Ottoniel (1990) sobre la situación y el perfil de los maestros formados durante la decada de los ochenta, como los realizados posteriormente por Reimers (1995) con carácter de diagnóstico, señalan aspectos relacionados con el desempeño de los maestros y que llevan a una clarividencia, por parte de los propulsores de la reforma, sobre la necesidad de establecer mecanismos para la evaluación y calificación de la acción docente.

El estudio de Ottoniel recoge información acerca de las características profesionales y humanas de los maestros de las tres regiones del país (occidental, central y oriental) porque se considera que éstas son un elemento determinante, entre otros, para el mejoramiento de la calidad de la educación. Esta información permite establecer un perfil real y plantear los problemas cruciales a los que debe atender la evaluación de los docentes.

El análisis de los datos recogidos establece que la mayoría de los maestros, cuando egresan, son enviados a trabajar con los primeros grados lo cual dificulta el desarrollo de su experiencia docente, sobre todo si se tiene en cuenta que los mismos docentes son conscientes de las fuertes carencias que se dan en ellos referidas al dominio ortográfico, a los pasos a seguir para desarrollar un contenido programático y al conocimiento de métodos, lo que lleva a improvisar y a no planificar el trabajo del aula. A pesar de ello, los maestros están satisfechos con su profesión, excepto en lo salarial, porque les permite crear vínculos con los alumnos, sus familias y sus compañeros de trabajo.

El estudio revela una situación crítica :

- en el dominio de conocimientos, actitudes y prácticas de los docentes; el estudio

pone en evidencia que no hay una coherencia entre la práctica y la normativa existente para la incorporación de maestros al ejercicio docente;
- en el desempeño del maestro en el aula se deja mucho a la improvisación y más que una tarea de proponer situaciones creativas para el trabajo relacionadas con los intereses y necesidades de los alumnos, el maestro se dedica a repetir lo que dice un programa o texto y a hacer mediciones finales que a lo mucho señalan cúanto ha retenido en su memoria el alumno;
- las condiciones laborales de los maestros no ofrecen incentivos que los animen a perfeccionarse y a superar los vacíos que, como profesionales, enfrentan.

Todo ello lleva a los estudiosos a hacer las siguientes recomendaciones:

- que se dé cumplimiento a los acuerdos y a las leyes relacionados con el nombramiento de maestros;
- que a los recién egresados se les asigne un trabajo en el segundo ciclo ya que los maestros comienzan sin tener un dominio para abordar el proceso de enseñanza en los primeros grados;
- que la capacitación, actualización y buen desempeño de su función otorgue a los maestros incentivos apropiados tendientes a la superación escalafonaria y a mejores ingresos.

Por otra parte, el estudio de Reimers sienta los precedentes para un diagnóstico que hará posible encarar la educación como un proyecto nacional a través del cual se entiende la necesidad de invertir en educación y capacitación de recursos humanos de manera que se pueda iniciar una transformación estructural del sistema educativo. Contiene sugerencias que reflejan la necesidad de dotar al sistema de mecanismos propios y eficientes para la evaluación del trabajo de los docentes.

Insiste en que el nivel básico debe continuar recibiendo atención prioritaria para constituir la base de la pirámide educativa y, especialmente, para desarrollar un sistema de formación docente para maestros de primaria con atención a la formación inicial y a la capacitación en servicio. Se recomienda crear y/o fortalecer «talleres de formación para maestros en servicio» y promover "las reuniones de los maestros para intercambiar experiencias" pues "probando y equivocándose podrían ir mejorando permanentemente su trabajo en el aula". ¿Qué es esto sino un tipo de evaluación que debe hacerse sobre la acción docente?

Por otra parte, se reconoce que en el país "no existe un mecanismo específico que permita evaluar el rendimiento del sistema en su conjunto" suponiéndose y esperándose que esta función sea cumplida por la dirección de supervisión que por lo demás bien se sabe que está sobrecargada de funciones administrativas. Además, no cuenta con los elementos técnicos necesarios para ello de manera que el aparato ministerial no logra animar a los docentes ni controlar el cumplimiento de sus funciones.

Es cierto que existe normativa sobre el funcionamiento y control de la acción de los docentes, pero ésta se encuentra dispersa en varias leyes y reglamentos. En las mismas se observa una tendencia protectora hacia la función del docente expresada en la insistencia en la estabilidad laboral, los derechos como garantía a la ocupación plena y el régimen de licencias.

El sistema de escalafón contempla incentivos que premian la antigüedad y el tipo de formación inicial pero no la clase de desempeño que realiza, lo que contribuye aún más a la desmotivación y al mantenimiento en una permanente inercia.

La Ley General de Educación (art. 112, letra e) establece como un derecho de los educadores el que gocen de ascenso sobre la base de méritos y aptitudes; pero ¿cómo interpretar los méritos y las aptitudes? Mientras que es claro que el ascenso de categoría por tiempo de servicio es automático, el Reglamento General de Educación Media (arts. 121 y 122) amplía lo que se puede entender por méritos al especificar: trabajos de investigación pedagógica y estudios sobre problemas de educación o redacción de libros de textos. Sin embargo, estos méritos serán apreciados por una comisión nombrada *ad hoc*, lo que no es ninguna garantía de objetividad. De igual manera, tampoco se define lo que se entiende por «trabajador más eficiente de la dependencia» cuando ocurre una vacante.

Teniendo en cuenta la complejidad de todo lo señalado en estos estudios y a partir de la información entregada por los mismos, parece interesante resaltar el aporte que sobre la problemática de la evaluación docente hace la Propuesta de la Comisión Nacional de Educación, Ciencia y Desarrollo (1995) cuando sugiere la creación de un "sistema confiable de evaluación permanente del desempeño docente". Dicho documento contiene, además, un conjunto de distinciones importantes para que el Ministerio pueda llevar adelante la creación de dicho sistema. Tales distinciones se refieren a que:

- el sistema esté orientado a incrementar la eficiencia y eficacia de maestros, directores y subdirectores;
- la información que se entregue sirva para tomar decisiones relativas a la renovación de contratos, reubicación y otorgamiento de bonificaciones vinculadas al mérito;
- la función de evaluar el desempeño docente sea asumida por el director de la escuela;
- se realice al menos una vez al año y haya la misma participación de los alumnos y padres de familia;
- el Ministerio no esté ausente y ejerza una función de auditoría de manera selectiva o aleatoria;
- es conveniente elaborar una nueva Ley de Escalafón del Magisterio Nacional para poder definir los criterios y propósitos, especialmente en lo relativo a la clasificación de los docentes, a las causas de los ascensos que deben estar centradas en el desempeño y a los sueldos y sobresueldos que deben basarse en

principios de equidad y justicia.

RECOMENDACIONES

Este análisis demuestra la importancia de la evaluación del desempeño docente en los procesos de reforma que se están ejecutando en la región. La calidad de los aprendizajes está sujeta a varios factores, uno de los cuales es justamente la acción docente.

Dada la importancia de transformar y mejorar los sistemas educativos y las prácticas que en ellos se viven, se hace urgente enfrentar, no sólo por parte de las autoridades educativas, sino también por todos los agentes educativos y, de manera muy especial por los docentes, que se tomen medidas para la transformación y el mejoramiento de este aspecto si se quiere avanzar en la búsqueda de la calidad educativa.

Pero también, dada la heterogeneidad de los países de la región, las situaciones y características que en cada lugar hay que enfrentar, más que tomar decisiones para homogeneizar los sistemas educativos, lo que interesa es apropiarse de criterios que garanticen la calidad de los procesos.

¿Son igualmente válidas para todos los países de la región estas recomendaciones? ¿No nos encontramos ante una realidad que es mucho más heterogénea de lo que se cree y que requiere de recomendaciones diferenciadas según la situación específica de cada región? ¿No es mejor pensar en criterios para una acción política en lo educativo que en recomendaciones prácticas respecto a lo que hay que hacer? El presente trabajo concluye sugiriendo algunas recomendaciones en tal sentido.

Recomendaciones sobre investigación

1) Es necesario tener un mayor conocimiento sobre la práctica de la Evaluación del Desempeño Docente desde diferentes perspectivas: la normativa legal, la práctica de la evaluación de los docentes en las instituciones escolares y las percepciones y reacciones de los gremios frente al tema.

Sería especialmente útil contar con más estudios sobre los enfoques, procedimientos y estrategias de evaluación del desempeño docente en experiencias y programas de innovación.

2) Sería de suma importancia conocer cuáles son los verdaderos motivos por los que el tema de la evaluación docente produce tanta resistencia en el gremio, así como si estas resistencias se producen de igual manera en aquellas situaciones en las que la evaluación es hecha por agentes externos al centro educativo, que cuando es realizada por agentes del mismo centro educativo. Si bien hay una práctica compartida y se habla de resistencia de los maestros a las evaluaciones por el uso que se hace con el resultado de las mismas, no hay información escrita ni

sistematizada sobre la experiencia acumulada.

3) Son pocas las experiencias que se conocen en este campo. No se sabe con seguridad si es que no existen o que no se difunden sus resultados. Si prima lo segundo es que aún nos encontramos en una concepción fiscalizadora que puede calzar con esquemas educativos de gran verticalismo, pero nada coherentes con los enfoques que las reformas quieren promover en la región. Este es otro campo importante para la investigación.

4) Sería importante promover experiencias de utilización de evaluaciones docentes con fines de toma de decisiones en materia de salarios, designaciones, cargos y responsabilidades creadas y asumidas en la misma experiencia.

5) Igualmente ayudaría conocer experiencias de centros educativos que han determinado sus propios criterios de evaluación docente a partir de su propio proyecto educativo y la información aportada por los expertos y la investigación.

Recomendaciones sobre política

1) Los sistemas educativos deben estar dotados de un sistema de evaluación docente que determine la finalidad con que se evalúa a los docentes, los criterios para evaluar su desempeño y los procedimientos de la evaluación. Es impensable realizar una evaluación con el propósito de escalar puntaje, gratificar o castigar. La evaluación de docentes, tanto de los nuevos como de los que ya están en servicio, debe orientarse hacia la determinación y el fortalecimiento de estrategias para la mejora de la calidad educativa.

2) Los sistemas de evaluación ya existentes deben estar abiertos a la inclusión de nuevos criterios que surgen desde la práctica docente para dar un mayor protagonismo y participación a los mismos docentes en la evaluación de su desempeño. La evaluación sólo puede incidir en la transformación de la educación si se estudia con la participación de los directos encargados de la cotidianeidad de la escuela. Por lo tanto, es necesario que los diseños, las propuestas y hasta los procedimientos que se hagan sean a partir de la experiencia acumulada y reflexionada de los que están permanentemente vinculados al día a día de la realidad pedagógica y escolar.

3) Dentro de los procesos de descentralización que se desarrollan con éxito en la región, los Ministerios de Educación deben participar activamente en la creación de mecanismos eficientes para apoyar, de manera permanente y técnica, a cada uno de los sistemas educativos descentralizados en la planificación y elaboración de instrumentos adecuados y el análisis de los resultados.

4) Se hace necesario revisar los actuales mecanismos que pueden estar reflejando concepciones ya superadas sobre lo educativo y la evaluación.

5) Es urgente seguir promoviendo y difundiendo aquellas experiencias exitosas que

sobre evaluación del desempeño docente se hagan en los diferentes países de la región.

NOTAS

[1] Agradecimientos: A Blanca Otárola, Colombia, y Nora Basso, Argentina, por su colaboración para la búsqueda de información.

[2] En un encuentro de docentes para abordar el tema, una maestra señala: «Acá no existe. En la práctica la hace el director de manera subjetiva y en función de su simpatía. En otros casos se hace de manera más objetiva pero poco sistemática. Cuando hay problemas de disciplina con algún docente o éste tiene muchos aplazados, el Director ejerce presión, «ese docente no va».

[3] Esto quedaría más tarde plasmado en la ley de la siguiente manera: "Para la Evaluación y Calificación de Servicios del personal comprendido en la Carrera Docente, se usarán los formularios que deberá confeccionar la Dirección General, de acuerdo con el Administrador General de Enseñanza, siguiéndose las indicaciones del Manual de Evaluación y Calificación de Servicios, que igualmente preparará la Dirección General" (Ley de Carrera Docente, cap.IX, art.75).

REFERENCIAS

Ahumada, P. (1992). *Evaluación de la eficiencia docente*. Chile: Universidad Católica de Valparaíso.

Arancibia, V. (1994). Características de los profesores efectivos en Chile y su impacto en el rendimiento escolar y autoconcepto académico. *PSYKHE*, Vol.3, No.2.

_____ (1987). *Estado del arte: manejo instruccional del profesor en la sala de clases en América Latina*. Projet BRIDGES. Boston: Harvard University.

Avalos, B. Y Haddad, W. (1981). *Reseña de la investigación sobre la efectividad de los maestros en el Africa, América Latina, las Filipinas, la India, Malasia, el Medio Oriente y Tailandia: síntesis de resultados*. Bogotá: CIID.

Cardemil, C. y otros. (1991). Factores que inciden en el mejoramiento de los aprendizajes en educación básica. *Cuadernos de Educación*, 208. Santiago de Chile: CIDE.

Carranza, R. (1992). *Modelo de evaluación para el desempeño docente*. Facultad de Educación de la Universidad de Costa Rica.

Connelly, F.M. (1989). *Evaluación de profesores: una revisión crítica para una práctica reflexiva supervisada*. Ponencia presentada al Seminario sobre Políticas de Remuneración y Formación de Profesores. Montevideo.

Ezpeleta, J. (1989). *Escuelas y Maestros.Condiciones del trabajo docente en la Argentina*. Santiago de Chile: UNESCO-OREALC.

Fallas, A. y otros. (1993). *Propuesta de modelo de evaluación del desempeño*

docente,con su respectivo manual,normas,procedimientos e instrumentos: caso del sistema educativo Saint Clare. Facultad de Educación Universidad de Costa Rica.

Filp, J. (1994). Todos los niños aprenden.Evaluaciones del P-900. En *Cooperación Internacional y Desarrollo de la Educación*. Chile: ASDI,AECI,CIDE.

Galo de Lara, C. (1990). *El maestro como orientador del aprendizaje.Evaluación del desempeño*. Guatemala: CINDEG.

García-Huidobro, J.E. (1994). El Programa de las 900 Escuelas. En *Cooperación Internacional y Desarrollo de la Educación*. Chile: ASDI,AECI,CIDE.

Garro, G. (1988). El pago por méritos y los problemas de la supervisión y la evaluación de los educadores. En *Revista de Educación de la Universidad de Costa Rica*. Vol.12, No.1. San José de Costa Rica.

González Soler, A. (1980). El perfil del profesor eficaz como base para la evaluación de los programas de formación del profesorado: problemas y perspectivas. En *La investigación pedagógica y la formación de profesores*. Madrid: C.S.I.C.

Mejía, M. (1987). *Movimiento pedagógico. Una búsqueda plural de los educadores colombianos*. Bogotá: CINEP.

Namo de Mello, G. (1982). *Magisterio de1 Grau. Da Competencia Técnica al Compromiso político*. Sao Paulo: Autores Asociados.

Núñez, I. (1990). *Las organizaciones de los docentes en las políticas y problemas de la educación*. Santiago de Chile: UNESCO-REDUC

_____. (1996). *El trabajo docente: condiciones y regulaciones. Documento mimeografiado*. Santiago de Chile: Ministerio de Educación.

Ordóñez, C. y otros. (1996). *Propuesta de organización del Sistema de Selección para Ingreso a la Carrera Docente*. Santafé de Bogotá: ICFES.

Ottoniel, S. (1990). *Estudio del perfil real y de la satisfacción profesional del maestro salvadoreño, formado durante la década de 1980-1989*. El Salvador: ED-UCA.

Popham, J. (1980). Problemas y técnicas de la evaluación educativa. ANAYA. Madrid.

Reimers, F. (1995). *La educación en El Salvador de cara al siglo XXI. Desafíos y oportunidades*. San Salvador: UCA Editores.

Rojas, C.y Castillo, Z. (1987). *Evaluación del programa Escuela Nueva en Colombia*. Bogotá: Instituto SER de Investigaciones.

Rossenfeld, M. (1983). "Valoración de la actuación de los docentes: un panorama." *ACADEMIA, 5-6*. Academia Superior de Ciencias Pedagógicas de Santiago de Chile.

Rubilar, M. y Cuevas, S. (1996). La supervisión pedagógica, una variable de apoyo al proceso de modernización de la educación. Resultados de una experiencia. *Revista de Pedagogía*, marzo 1996. FIDE. Santiago de Chile.

Schön, D. (1987). *Educating the reflective practitioner*. Londres: Jossey-Bass.

Universidad Nacional de Colombia (1996). *Evaluación de Docentes, Proyecto UN-MEN*. Santafé de Bogotá.

Villa Sánchez, A. (1985). La evaluación del profesor: perspectivas y resultados. *Revista de Educación*, 277. Madrid.

Parte IV

EVALUACIÓN DE LA ORGANIZACIÓN EDUCATIVA

CAPÍTULO 9

EVALUACIÓN DE LAS ESCUELAS

William J. Webster y Robert L. Mendro

La cuarta parte de este libro está dedicada a un tema que cada día cobra mayor importancia, dada la dirección de las nuevas políticas de la educación y el inexorable movimiento hacia la mayor participación de los padres de familia y las comunidades en el gobierno escolar: la evaluación de la escuela. Esta parte se inicia con un capítulo sobre la evaluación de la escuela, que se inscribe en la tradición de la investigación social y la evaluación y presenta en detalle la utilización de los componentes absolutos y de valor agregado en la evaluación de las escuelas, los distritos escolares y el sistema escolar estatal o nacional. Los autores ilustran sus contribuciones teóricas y metodológicas a través del caso de las escuelas públicas de Dallas en los Estados Unidos, donde se ha aplicado la evaluación de valor agregado con el fín de mejorar la educación y de facilitar a los padres de familia, profesores y demás miembros de la comunidad escolar un conocimiento sobre el progreso de cada componente del sistema.

INTRODUCCIÓN

Esta monografía presenta un marco de referencia para el desarrollo de opciones de política para evaluar el desempeño individual de las escuelas dentro del contexto de una reforma educativa nacional que incrementa la descentralización y autonomía local. Se presentará primero una breve revisión de los modelos de evaluación existentes. A continuación se hará una breve reseña de los tres niveles que se utilizan para el establecimiento de responsabilidades dentro del sistema. El primer nivel es el de la escuela. Cada escuela debe tener a su disposición datos relacionados con resultados educativos importantes que ameriten medición y a partir de los cuales se desarrollen los estándares. Este trabajo discute la relación entre estándares externos e internos—propios de la escuela—y sus respectivas mediciones, y establece además la necesidad de evaluar su contexto, insumo, proceso y producto. El segundo nivel es el distrital o nacional. Las agencias nacionales o distritales fijan los niveles deseados de objetivos para el establecimiento de las responsabilidades

correspondientes y especifican la naturaleza de la ayuda que se debe brindar a cada escuela y los vehículos a utilizarse para cumplir con esta ayuda. El tercer nivel corresponde al desarrollo de mediciones que no presenten un sesgo y que puedan ser usadas en la evaluación de valor agregado. Estas mediciones tienen en cuenta antecedentes importantes de la escuela y del estudiante, particularmente su rendimiento académico anterior, y se lleva a cabo con el fín de disponer de información adecuada sobre el grado de eficacia de la escuela que beneficie a todos y cada uno de sus estudiantes. Este enfoque permitirá diseñar sistemas a niveles nacional y de la escuela, que examinen tanto los componentes absolutos como los de valor agregado, de tal manera que se obtenga una medición comprensiva de los efectos y de la eficacia de la escuela.

A medida que las escuelas adquieren mayor autonomía y control de sus recursos, la necesidad de establecer responsabilidades sobre su operación y logros se hace más urgente. La gerencia basada en la autonomía de cada escuela conlleva una alta dosis de responsabilidad local para garantizar que los estudiantes reciban una educación adecuada. El proceso por medio del cual se establecen responsabilidades es la piedra angular sobre la cual se construye un sistema de gerencia escolar descentralizado. El sistema de evaluación debe difundir información adecuada a los diferentes usuarios de los servicios escolares. Esta información sirve para tomar decisiones locales, supervisar aspectos más generales del sistema y establecer responsabilidades. Para los propósitos de este capítulo, un estudio de evaluación es aquél que se diseña para ayudar a las audiencias o personas interesadas a juzgar el valor de un objeto educativo y mejorarlo. Esto sirve tanto para la toma de decisiones, como para la identificación de responsabilidades.

CONCEPTUALIZACIONES ALTERNATIVAS DE EVALUACIÓN

Stufflebeam y Webster (1980) caracterizaron trece enfoques diferentes de evaluación. Estos enfoques fueron agrupados en las siguientes categorías: estudios de orientación política (llamados "pseudo evaluación"), estudios orientados a responder preguntas (llamados "cuasi evaluación") y estudios orientados a juzgar el valor de un objeto (llamados "de verdadera evaluación"). En los cuadros 1 a 3 se presenta una revisión concisa de los tres tipos diferentes de estudios y los trece enfoques diferentes de evaluación. Estos cuadros esbozan los tipos de estudio, los organizadores conceptuales más representativos de cada tipo de estudio, las fuentes de los interrogantes y las preguntas principales que se plantea cada uno de ellos y los métodos más comúnmente utilizados. También se incluye el nombre de los pioneros de dos tipos de estudio: los orientados a preguntas y los orientados a valores.

La mayor parte del trabajo inicial de evaluación educativa tuvo lugar entre 1960 y 1970. El trabajo más reciente se ha concentrado en el desarrollo de metodologías apropiadas para enriquecer la práctica de la evaluación, más que en los desarrollos teóricos *per se*. Stufflebeam, un prolífico escritor, ha puesto más su atención en desarrollar los estándares para la evaluación educativa que en ampliar el modelo que

Cuadro 1.a Tipos de estudios según su orientación

Cuadro 1.b Tipos de estudios según su orientación

Cuadro 1.c Tipos de estudios según su orientación

propuso. Este modelo, Contexto-Insumo-Proceso-Producto (CIPP), ha sido el de mayor influencia en las escuelas de los Estados Unidos (Webster,1988). Otros trabajos que ofrecen presentaciones y revisiones de varios modelos de evaluación, son: *Evaluation Models (Modelos de evaluación)* de Madaus, Scriven y Stufflebeam (1983) y *Meta-Evaluation of School Evaluation Models (Metaevaluación de los modelos de evaluación) de* Gallegos (1994). La primera de estas publicaciones incluye ensayos escritos de algunos de los más importantes teóricos de la evaluación. El trabajo de Gallegos es citado ampliamente, puesto que describe 51 modelos diferentes en operación. Stufflebeam (1996) publicó posteriormente un excelente marco teórico para la evaluación de estudiantes, programas y personal.

Consideramos que los enfoques de evaluación más apropiados incluyen los estudios orientados a valores y, dentro de este grupo, los estudios que sirven de base a las decisiones. Estamos de acuerdo con este punto de vista porque los tipos de estudio orientados hacia decisiones se encaminan no sólo al establecimiento de responsabilidades, sino que ofrecen información útil para la toma de decisiones y el mejoramiento de los servicios sociales. El establecimiento de responsabilidades sin información para el mejoramiento no es muy útil, si se trata de encontrar caminos para mejorar la calidad del sistema que se está examinando. Creemos también que el diseño de cualquier evaluación debería apoyarse en ciertos requisitos o condiciones, conocidos como "estándares de evaluación". Los estándares de evaluación se pueden agrupar en cuatro conjuntos: estándares de utilidad, factibilidad, propiedad y precisión. A este respecto se pueden consultar los Estándares de evaluación de programas (Program Evaluation Standards) desarrollados por el Comité para el Desarrollo de Estándares para la Evaluación de la Educación (Joint Committee on Standards for Educational Evaluation) en 1994. Si una evaluación no cumple con tales estándares, quienes toman decisiones sobre las escuelas se ven forzados a hacerlo basados en información inexacta, inválida, incompleta o incomprensible.

CONSIDERACIONES PARA EL DESARROLLO DE UN MODELO DE EVALUACIÓN DE LA ESCUELA

Para el desarrollo de un modelo de evaluación para las escuelas que cumpla los requisitos antes mencionados, los autores se han inspirado principalmente en el estudio de Stufflebeam y otros (1971), Scriven (1967), Stake (1967) y Provus (1971) en conjunción con los avances realizados en el campo de la metodología para la medición del valor agregado, específicamente el modelo linear jerárquico (HLM, por sus siglas en inglés Hierarchical Linear Model), como lo describen Bryk y Raudenbush (1992), Bock (1989) y Goldstein (1987). A continuación se presenta una breve descripción de cada uno de estos componentes. Un lector cuidadoso ya habrá notado que se trata de un tipo de estudio clasificado anteriormente como orientado a la toma de decisiones.

Modelos de evaluación

Probablemente el modelo de evaluación más completo ha sido el desarrollado por

Stufflebeam y otros (1971), llamado el modelo CIPP. La evaluación es definida como el proceso de esbozar, obtener y proporcionar información útil para juzgar decisiones alternativas. El modelo identifíca cuatro tipos importantes de evaluación: la evaluación de contexto, que alimenta las decisiones de planificación; la evaluación de insumos, que alimenta las decisiones de programación; la evaluación de proceso, que alimenta las decisiones de ejecución y la evaluación de producto, que alimenta las decisiones de reciclaje.

En síntesis, la evaluación de contexto sienta las bases para la determinación de objetivos educativos y la definición de ambientes importantes para la descripción de las condiciones deseables y reales del ambiente, y para la identificación de necesidades insatisfechas y el diagnóstico de los problemas que dificultan la satisfacción de las necesidades. La evaluación de insumo evalúa las capacidades de las agencias responsables e identifíca estrategias para alcanzar los objetivos propuestos a través de la evaluación de contexto, así como también presenta sugerencias encaminadas a ejecutar las estrategias seleccionadas. Una vez que se ha escogido una estrategia, el proceso de evaluación proporciona información de retorno a las personas encargadas de la ejecución de los planes y procedimientos para predecir o detectar fallas en los diseños. De esta manera se pueden garantizar los ajustes en el proceso. Finalmente, la evaluación de producto ofrece estimaciones sobre los efectos de los programas educativos mientras se desarrollan y cuando se terminan. Esto quiere decir que la evaluación de producto identifica los efectos de las estrategias seleccionadas a través de una evaluación de insumos, la cual a su vez satisface las necesidades identificadas por medio de la evaluación de contexto. Esta evaluación se completa a la luz de la información obtenida durante la evaluación de proceso.

Scriven (1967) conceptualizó un marco de referencia de evaluación muy claro y ampliamente aceptado. No es tan amplio como el Modelo CIPP y se concentra principalmente en la dimensión proceso-producto del modelo propuesto por Stufflebeam. De acuerdo con Scriven, la meta principal de la evaluación es la de emitir, con suficiente credibilidad, juicios relacionados con el mérito y el valor de un programa educativo. Dentro de la discusión de métodos para alcanzar esta meta, Scriven introduce el concepto de evaluación formativa y sumativa.

El foco de la evaluación formativa es el mejoramiento del respectivo programa. De ahí que la evaluación formativa busca ofrecer información de retorno a su personal directivo, con el propósito de mejorar dicho programa mientras se encuentra en etapa de desarrollo. En el CIPP original, los datos sobre el proceso y los productos parciales proporcionan, al personal del programa, la información sobre la evaluación formativa.

El foco de la evaluación sumativa es la determinación del valor último de un programa o proyecto. Este tipo de evaluación debe ser ejecutada en la etapa en que el programa ha alcanzado estabilidad. Los datos sumativos ayudan a la toma de decisiones fundamentales sobre el futuro del programa. Como resultado de la

información suministrada por la evaluación sumativa, un programa se puede acabar, reestructurar, continuar o ampliar. En el CIPP original, la información de la evaluación del producto final, interpretado en relación con los datos del contexto, insumos y procesos, se emplea para deducir conclusiones sumativas acerca del mérito o valor de un programa educativo y para fundamentar las decisiones sobre su continuación o cambio.

Stake (1967) señaló que la evaluación tiene que ver con tres clases de condiciones: antecedentes, transacciones y resultados. Los antecedentes se definen como aquellas condiciones que existen antes de iniciarse el programa, por ejemplo, el contexto educativo. Las transacciones son las interacciones entre los estudiantes, profesores y materiales. Los resultados se definen como los productos que se espera obtener de las transacciones.

Stake sugiere tres clases de actividades para la realización de una evaluación. La primera incluye la ayuda que se le brinda al personal del programa para generar una fundamentación clara del programa o proyecto. La segunda consiste en la generación de datos descriptivos. Los datos descriptivos incluyen planteamientos relacionados con antecedentes, transacciones y resultados actuales y propuestos. Por lo tanto, se debe hacer una revisión acerca de la congruencia entre antecedentes, transacciones y resultados planeados y observados. Stake también sugiere efectuar un examen de contingencia de los antecedentes, las transacciones y los resultados propuestos (análisis de contingencia lógica), así como de antecedentes, las transacciones y los resultados observados. El análisis de contingencia lógica es similar a la evaluación de insumo en el modelo CIPP, mientras que el análisis de los datos observados intenta identificar relaciones de causa y efecto entre antecedentes, transacciones y resultados.

La tercera clase de actividades propuestas es la formulación de juicios en torno del valor de los programas educativos. Stake sugiere que tales juicios se hagan considerando los criterios absolutos y relativos. En otras palabras, los programas deben ser evaluados en términos del grado en que alcanzan metas absolutas y de su nivel de desempe o relativo, cuando se comparan con otros programas que tienen objetivos similares.

Provus (1971) propuso que todos los proyectos pasaran por las etapas de diseño, instalación, proceso y producto. Durante cada una de estas etapas el evaluador debe delinear, con los otros miembros del grupo, un conjunto de estándares que se puedan utilizar como base de comparación del rendimiento del programa. Es función del evaluador hacer comparaciones entre los estándares y el rendimiento, identificar las discrepancias en cada etapa e informar sobre dichas discrepancias al director del proyecto quien tiene la opción de terminar el programa, modificarlo, o avanzar hacia otra etapa. El producto de la etapa de diseño es un conjunto de estándares utilizados para juzgar los efectos de los esfuerzos del programa en cada una de las tres etapas descritas. En cada una de las fases, el objeto de la evaluación es ofrecer datos útiles para la toma de decisiones sobre el mejoramiento del programa.

El foco principal de las cuatro perspectivas de la evaluación mencionadas anteriormente es la evaluación de programas. Su aplicación al caso de la evaluación de la escuela es directa, como se observará a lo largo de este documento.

Utilidad de los resultados no ajustados versus los resultados de valor agregado en la evaluación de la escuela

Hasta este momento se han presentado varios sistemas de evaluación de la escuela, se han ilustrado los diversos modelos de trabajo que han surgido y se han discutido, a partir de tales sistemas, las formas en que el contexto, el insumo y los procesos se relacionan entre sí y con los resultados. Pero, sólo se han considerado los resultados no ajustados. Algunos resultados importantes de la educación deben ser no ajustados. Sin embargo, el uso de resultados no ajustados—sin tener en cuenta sus característícas—puede conducir a evaluaciones de la escuela sesgadas o equivocadas.

El propósito de esta sección es definir y discutir el uso de los resultados de valor agregado de la evaluación de la escuela en contraste con el uso de los resultados no ajustados. Esta sección ilustra los criterios con relación a los cuales se deben medir todos los resultados, sean éstos no ajustados o de valor agregado. Además presentan algunos ejemplos de errores de interpretación del efecto de las escuelas cuando se emplean los resultados no ajustados.

Para explicar mejor este punto, comencemos con dos ejemplos de resultados no ajustados y de las metas establecidas sobre la base de tales resultados. Supongamos, en el primer caso, que una escuela tiene una tasa de deserción escolar del 15 por ciento y se ha establecido una meta de reducción de esta cifra al 13 por ciento. En el segundo caso, imaginemos que cada una de las escuelas de un sistema tiene al menos el 50 por ciento de sus estudiantes leyendo al nivel adecuado a su respectivo curso y que la meta del sistema es que cada escuela logre tener 60 por ciento de sus estudiantes leyendo al nivel correspondiente. A primera vista estos planes parecen hacer un uso realista de los resultados no ajustados y establecer metas razonables.

Consideremos estos dos ejemplos más detenidamente. Supongamos, en el primer ejemplo, que la población de jóvenes de donde salen los estudiantes de nuestra hipotética escuela, tiene una historia general de una tasa de deserción escolar que oscila entre el 9 y el 11 por ciento. Ahora bien, un resultado no ajustado del 15 por ciento no es muy deseable y una meta del 13 por ciento es demasiado alta. Una meta más apropiada para la escuela sería la reducción de la tasa de deserción al 12 por ciento en el primer año y por debajo del 11 por ciento a partir de entonces. En el segundo ejemplo, supongamos que en los dos años anteriores cada escuela haya tenido en algún momento 60 por ciento de sus estudiantes leyendo al nivel de su respectivo curso o grado. Una meta no ajustada para cada escuela podría ser la de lograr que por lo menos el 65 por ciento de sus estudiantes leyeran al nivel que les corresponde.

El lector habrá notado el cuidado que es necesario tener al interpretar estos ejemplos. Las metas y los resultados están condicionados por los niveles de desempeño en el pasado de la población estudiantil a la cual nos referimos. Al revisar de nuevo los dos ejemplos con condiciones subyacentes diferentes, los mismos resultados no ajustados y sus respectivas metas se tornan inapropiados por diversas razones. En el primer ejemplo, la escuela, con una población dada de estudiantes, ha mantenido sus tasas de deserción en un 15 por ciento. Si suponemos que las tasas de deserción de la escuela oscilan entre el 18 y el 20 por ciento y que ninguna otra escuela, con poblaciones similares ha podido reducir estas tasas a menos del 17 por ciento, la meta del 13 por ciento puede resultar poco realista por diferentes razones. La tasa del 15 por ciento puede constituir un ejemplo de la mejor práctica en curso. Con toda probabilidad, una tasa no ajustada del 15 por ciento constituye un excelente resultado. La escuela puede ser elogiada por mantener esta tasa y debe compartir, con las otras escuelas, las técnicas utilizadas. La tasa no ajustada del 15 por ciento se considera por el momento un excelente resultado, hasta cuando surjan otras técnicas más efectivas que permitan reducir esta tasa por debajo del 15 por ciento.

En el segundo ejemplo, la meta del sistema fue que el 60 por ciento de los estudiantes pudieran leer al nivel de su respectivo curso en cada escuela. Supongamos ahora que este es el primer año en que todas las escuelas lograron que el 50 por ciento de sus estudiantes pudieran leer al nivel de su curso. Imaginemos también que en algunas escuelas el 75 por ciento de los estudiantes nunca ha estado por debajo del nivel de lectura de su respectivo curso. La meta del sistema de 60 por ciento puede ser más apropiadamente formulada como 55 por ciento para algunas escuelas y 80 por ciento para otras.

En ambos ejemplos el contexto juega obviamente un papel trascendental. Por tanto, las metas y los resultados deben ser considerados a la luz de cada contexto. Como Glass (1978) señalaba, los resultados, las metas, o los estándares desarrollados a partir de las metas deben estar relacionados con el desempeño actual de los grupos específicos de estudiantes a los que se refiere una evaluación. Este ejemplo también ilustra uno de los mayores problemas asociados con la utilización de los resultados sin ajustar. Con frecuencia se establecen metas absolutas basadas en tales resultados sin pensar si se pueden alcanzar. Webster y Mendro (1995) estudiaron ampliamente este tema y concluyeron que es posible establecer metas alcanzables basadas en resultados no ajustados. El problema radica en que a la administración pública o a las directivas les gustaría ver progresos masivos y rápidos y típicamente consideran que las metas construidas cuidadosamente y ejecutadas de manera incremental presentan un problema de "bajas expectativas". Las metas extremadamente altas y poco realistas y con escasa probabilidad de éxito son por lo general más satisfactorias para quienes están fuera de la escuela. Hemos constatado un problema adicional. Muy pocos educadores están en capacidad de hacer una transición directa de las metas medidas en términos de resultados académicos a los planes de acción conducentes al mejoramiento de la escuela. Este es un problema que se extiende a todos los esfuerzos de mejoramiento de la educación y será discutido más adelante

en este capítulo.

Más allá del problema del establecimiento de metas utilizando datos no ajustados, sin embargo, existe una pregunta sin responder sobre cómo evaluar el progreso educativo de una manera justa y precisa. ¿Cómo podemos determinar contextos apropiados para evaluar resultados educativos al interior de la escuela? ¿Cómo podemos determinar si los resultados alcanzados constituyen en realidad una mejora o si son simplemente el producto del progreso normal de una población definida?

El rápido desarrollo del área de evaluación del valor agregado de los resultados educativos nos puede ayudar a responder a estas y a otras preguntas similares. Con sistemas de valor agregado se pueden mantener constantes las condiciones fuera del control de la escuela, en todas las escuelas de un grupo. Los efectos obtenidos por cada escuela se miden con un mismo patrón o métrica común y se comparan sus resultados. En esencia, se establece una línea de base para todas las escuelas de un grupo; el elemento crucial consiste en saber si, en relación con otras escuelas, una escuela en particular ha tenido un efecto positivo o negativo en sus estudiantes. El interrogante se reduce a saber si la escuela añadió o sustrajo valor a la línea de base. Esta línea de base común ayuda a responder a las preguntas sobre el mejoramiento realizado a partir de un desempeño típico.

Se pueden utilizar una gran variedad de metodologías para desarrollar estos sistemas, las cuales ofrecen alternativas preferibles a los sistemas basados en resultados no ajustados. Aunque discutiremos un sistema que ha sido cuidadosamente investigado y ajustado para eliminar numerosos sesgos pequeños, la mayoría de los sistemas de valor agregado, basados en las técnicas de regresión, constituye una alternativa mucho mejor que los sistemas no ajustados. El propósito de estos sistemas es eliminar los factores conocidos que afectan los resultados de la escuela pero que ésta no puede controlar. Entonces sí se pueden identificar los avances. Adicionalmente, cuando estos sistemas están diseñados apropiadamente, se constituyen en la mejor alternativa para ajustar, a nivel de los estudiantes, los efectos de las variables que no se han incluido explícitamente entre los factores conocidos. En otras palabras, ellos ayudan a controlar hasta donde es posible, todos aquellos factores que no están bajo el control directo de las escuelas.

Entre los factores conocidos que afectan el progreso de la escuela pero que están fuera de su control, se pueden mencionar los siguientes: el nivel socioeconómico, el género, la raza, la utilización del lenguaje y las habilidades individuales de los estudiantes que ingresan. Los criterios necesarios para juzgar los sistemas de evaluación de las escuelas son los cursos de relación existentes entre tales factores y las medidas de eficiencia empleadas. No obstante, esto no sucede con frecuencia, según pudimos inferir a partir de una extensa búsqueda de la literatura.

Los sistemas que utilizan resultados no ajustados o programas de pruebas como base para la evaluación, demuestran también que sus resultados están altamente

correlacionados con los factores antes delineados. Como anotan Jaeger (1992), Webster y otros (1995), estos tipos de sistemas están sesgados a favor de las escuelas que tienen un alto porcentaje de estudiantes blancos de nivel socioeconómico alto. Una comparación entre las escuelas de Dallas, ordenadas de acuerdo con el sistema de valor agregado y las mismas escuelas clasificadas bajo un sistema de acreditación no ajustado, mostraron claramente que las escuelas efectivas (según la evaluación de valor agregado) que también arrojaron buenos resultados en el sistema no ajustado tienen un alto porcentaje de estudiantes blancos y de aquéllos provenientes de nivel socioecómico alto. Esto también indica que las escuelas que son muy efectivas con sus poblaciones, pero que tienen un alto porcentaje de sus estudiantes menos favorecidos económicamente, rinden mucho menos en sistemas sin ajustar. Finalmente, este mismo estudió mostró que las correlaciones entre las clasificaciones de la eficacia de las escuelas y los factores demográficos mencionados, es demasiado alta en sistemas no ajustados, pues llega a alcanzar el 0,90 a nivel de los alumnos y el 0,65 a nivel de la escuela (Webster y otros, 1995).

La esencia del argumento es que con resultados no ajustados las escuelas son clasificadas básicamente por el tipo de estudiantes que reciben y no por la clase de educación que imparten. El uso de resultados no ajustados para comparar las escuelas y los programas confunde las diferencias existentes en la población estudiantil y la forma como tales estudiantes fueron seleccionados en sus escuelas y programas, con la diferencia que estas escuelas y programas producen. Las escuelas y programas que atraen estudiantes con buenas calificaciones reciben un beneficio antes de que los estudiantes inicien sus clases. Las escuelas y programas que tienen que aceptar estudiantes con bajo rendimiento se ven obligadas a superar las dificultades iniciales antes de poder mostrar algún resultado.

Elementos necesarios para un modelo de evaluación de la escuela

Un sistema de evaluación escolar debe cumplir con una serie de requisitos si quiere proporcionar estimaciones útiles de los efectos de la escolarización. A continuación se presentan tales requisitos y se discuten luego con mayor detenimiento. Para resumir, un modelo de evaluación de la escuela debe incluir los nueve elementos siguientes:

1. Debe ser de valor agregado.
2. Debe incluir componentes para la evaluación del contexto, el insumo, el proceso y el producto.
3. Debe estar bajo el control de un grupo representativo de la comunidad escolar (incluyendo padres de familia, empresarios, etc.).
4. Debe incluir un amplio acuerdo para la medición de los resultados.
5. Debe estar basado en un análisis de la matrícula continua de los estudiantes en la escuela.
6. Debe incluir la obligación de examinar prácticamente a todos los estudiantes elegibles.
7. Debe incluir medidas previas de todas las variables de resultados, tanto a nivel de

los estudiantes como de la escuela (imparcialidad).
8. Debe controlar, a nivel de la escuela y de los estudiantes, las variables contextuales sobre las cuales las escuelas no tienen control (imparcialidad).
9. Debe proporcionar información para mejorar.

El primer requisito, que la evaluación debe ser de valor agregado, como ya se ha señalado, significa que las escuelas solamente son responsables por el progreso o ausencia de éste que ellas logran en los estudiantes que le han sido asignados. De ahí que es importante medir de alguna forma el progreso de los estudiantes. Este requisito al hacer una evaluación de las escuelas ayuda a limar las asperezas producidas por el uso exclusivo de estudios basados en objetivos, en programas de pruebas y sistemas de manejo de la información.

Esto parece sencillo, pero el concepto de valor agregado implica un valor agregado a *algo*. Con el objeto de poder determinar qué es lo que la escuela añade a la educación de los estudiantes, deberá hacerse una medición inicial que se pueda relacionar con los resultados alcanzados. Un sistema que intenta hacer desaparecer los efectos de las medidas relacionadas con los resultados sin suprimir los efectos de la habilidad de los estudiantes, arrojará altas correlaciones entre las medidas de habilidad inicial de los estudiantes y las de eficacia (Webster y otros, 1995). La inclusión de medidas tales como el nivel socioenómico, por sí misma, sin una medición previa de los logros, no controla substancialmente los efectos de logros anteriores del estudiante. La medición previa no tiene que ser —y ciertamente no siempre lo es— la misma medición que la de los resultados. El único requisito es que debe estar correlacionada con la medición de los resultados y estar relacionada directamente con la medición de destrezas similares o subyacentes.

En los Estados Unidos, los departamentos estatales de educación han desempeñado un papel de liderazgo en el proceso de ejecución de los sistemas de evaluación y asignación de responsabilidades. Cuarenta y seis de los cincuenta estados tienen sistemas que incluyen la determinación de responsabilidades por medio de algún tipo de evaluación. Veintisiete de ellos presentan informes a nivel de la escuela, del distrito o del estado; tres presentan informes solamente a nivel de la escuela; seis lo hacen a nivel tanto de la escuela como del distrito; siete presentan los informes a nivel del distrito y del estado; dos sólo lo hacen a nivel del estado; y uno se encuentra actualmente en etapa de desarrollo (Council of Chief State School Officer, 1995). La revisión de este sistema indica que sus diseñadores no están familiarizados con la literatura sobre sistemas de valor agregado, ya que sólo dos estados, Carolina del Sur (May, 1990) y Tennessee (Sanders and Horn, 1995) han utilizado apropiadamente la metodología estadística sobre valor agregado en la ejecución de estos sistemas. La mayoría de los otros estados tiende a evaluar a los estudiantes y no a las escuelas o los distritos, situación que puede causar más daño que beneficio por la desinformación sistemática en torno a las contribuciones de la escuela y los distritos a los logros académicos de los estudiantes. Con el propósito de evitar estos sesgos, varios estados han comenzado a utilizar técnicas no estadísticas de agrupación. Este enfoque tiene serias limitaciones cuando se presenta

consistentemente una varianza direccional relacionada con las características de agrupación dentro de los grupos.

Fennessey y Salganik (1983) propusieron un modelo para analizar la eficacia de los programas de instrucción dentro del contexto de los puntajes conseguidos. Estos autores desarrollaron un índice de puntaje ganado el cual, ajustado y trasformado en una escala conveniente (RAGS, o Rescaled and Adjusted Gain Score) permite igualar, en todos los grupos del programa, el efecto de factores tales como el sesgo neto agregado de la respuesta a la instrucción, la regresión a la media y otros. El supuesto básico en este enfoque es que cualquier grupo de estudiantes con resultados similares en el pre-test alcanzará tasas de aprendizaje similares y podrá ser sujeto del mismo grado de regresión a la media. Aunque el procedimiento RAGS es apropiado para la evaluación de programas, puede ser difícil de aplicar cuando se intenta determinar la eficacia relativa de las escuelas utilizando poblaciones estudiantiles diferentes.

El segundo requisito para estimar el valor agregado, que ha recibido amplia distribución y aceptación entre los investigadores de la educación, incluye la agregación de residuos de los modelos de regresión a nivel de los estudiantes (Aiken y West, 1991; Bano, 1985; Felter y Carlson, 1985; Kirst, 1986; Klitgard y Hall, 1993; McKenzie, 1983; Millman, 1981; Saka, 1989; Webster y Olson, 1988; Webster, Mendro y Almaguer, 1994). Estas técnicas incorporan un gran número de variables de insumo, proceso y producto en la ecuación y determinan el promedio de la desviación a partir de los resultados estimados para cada estudiante por cada escuela. Las escuelas son entonces clasificadas usando el promedio de la desviación. Algunas de las ventajas de los análisis de regresión múltiple en relación a otras técnicas para este caso, se refieren a la relativa simplicidad de su aplicación e interpretación, a su fortaleza y al hecho de que estos métodos generales de estructurar ecuaciones complejas de regresión e incluir combinaciones de varibles continuas y categóricas y sus interacciones son relativamente directos (Aiken y West, 1991; Cohen, 1968; Cohen y Cohen, 1983; Darlington, 1990).

Finalmente, el modelo lineal jerárquico (HLM, por sus siglas en inglés) proporciona estimados de ecuaciones lineales que explican los resultados de los miembros de un grupo, no sólo como una función de las características del grupo, sino también como una función de las características de sus miembros. El modelo lineal jerárquico comprende la predicción de resultados que están estructurados jerárquicamente dentro de grupos, los cuales pueden, a su vez, formar parte de grupos más grandes, estructurados también en forma jerárquica. Por esta razón es muy apropiado para utilizarlo en educación. Esta estructuración jerárquica dentro de las aulas de clase y de éstas dentro de la escuela, arroja diferente varianza a cada nivel para los factores medidos en ese nivel. Bryk y otros (1988b) señalan cuatro ventajas del modelo lineal jerárquico sobre los otros modelos generales de regresión lineal. Primero, puede explicar el rendimiento de los estudiantes y su avance como una función de las características nivel de la escuela o de la clase, tomando en cuenta la varianza de los resultados de los estudiantes dentro de las escuelas.

Segundo, puede ejemplificar los efectos de las características de los estudiantes tales—como el género, la raza o el nivel socioeconómico—sobre los resultados dentro de las escuelas o clases y explicar las diferencias en estos efectos entre las escuelas. Tercero, puede representar en un modelo las varianzas entre las escuelas y dentro de ellas al mismo tiempo y así predecir estimados más exactos de los resultados de aprendizaje del estudiante. Finalmente, puede producir mejores estimados de los predictores de los resultados de los estudiantes dentro de las escuelas y clases utilizando información "prestada" sobre estas relaciones en otras escuelas y clases. El modelo lineal jerárquico se ha estudiado ampliamente en la literatura bajo diferentes nombres y autores y diversas disciplinas (Bryk y Raudenbush, 1992; Dempster, Rubien y Tsutakawa, 1981; Elston y Grizzle, 1962; Goldstein, 1987; Henderson, 1984; Laird y Ware, 1982; Longford, 1987; Mason, Wong y Entwistle, 1984; Rosemberg, 1973).

El tercer requisito de un modelo de evaluación de la escuela es que las características del modelo deben estar bajo el control de un grupo representativo de la sociedad en que opera. El sistema debe facilitar una participación en los insumos a todas las personas involucradas. Estas personas deben formar parte de los cuerpos directivos encargados de tomar las decisiones finales sobre el sistema. Este grupo puede incluir a los padres de familia, miembros de la comunidad, representates del sector empresarial, administradores escolares, maestros, estudiantes, miembros de las directivas del sistema educativo y, en sistemas financiados por el gobierno, representantes nombrados por la institución correspondiente, estatal, nacional o local. Si se considera que aun el más simple de los sistemas presenta cierto nivel de complejidad matemática, es recomendable que un grupo de estadísticos sirva de asesor. Es especialmente útil que uno o más estadísticos formen parte del grupo de miembros de la comunidad o representantes de las empresas ya que, así, los otros miembros pueden recibir una confirmación independiente de las matemáticas involucradas, a través de uno de sus propios miembros. Esta confirmación independiente es fundamental para fortalecer la confianza del grupo en el sistema.

En Dallas, el organismo de gobierno establecido se llamó el Comité de Trabajo para el Establecimiento de Responsabilidades (Accountability Task Force) (Dallas Public Schools, 1996; Webster y otros, 1997). Este organismo incluye a padres de familia, miembros de la comunidad, representantes del sector productivo, maestros y directores y administradores que representan al superintendente de las escuelas. El Departamento de Investigación y Evaluación cumple la tarea de asesoría por medio de la oferta de servicios de análisis estadísticos al Comité de Trabajo.

El Comité de Trabajo para el Establecimiento de Responsabilidades toma las siguientes decisiones con respecto al sistema de valor agregado:

- Selección de resultados, variables dependientes y concomitantes en el sistema.
- Desarrollo y supervisión de normas y procedimientos del sistema.
- Atención a todas las solicitudes de procedimientos y resultados por escuela.
- Asesoría al Superintendente General y al Consejo de Educación sobre todas las

decisiones relacionadas con el sistema.

Las funciones legales y de asesoría del Comité de Trabajo para el Establecimiento de Responsabilidades son extremadamente importantes, puesto que el sistema de Dallas ofrece incentivos económicos a las escuelas más eficaces y el Consejo de Educación utiliza los resultados del sistema de evaluación para determinar el rendimiento de las escuelas. Ninguna escuela o director quiere que su destino en el sistema se decida sin poder contar con algún recurso cuando sea necesario.

El cuarto requisito que se considera esencial para evaluar cualquier escuela o sistema responsable es el de las variables de múltiples resultados. No es apropiado presentar informes de los resultados de pruebas referidas a normas o criterios y considerar estos informes como un sistema de evaluación con propósitos de asignación de responsabilidades. Las escuelas son responsables de múltiples resultados educativos. Además de las pruebas de resultados referidas a normas y criterios, otros resultados importantes de la escolarización serían, entre otros, los siguientes: muestras de trabajos escritos de los estudiantes y mediciones de desempeño, tasas de promoción, tasas de asistencia, tasas de graduación, tasas de deserción, matrícula en cursos avanzados y éxito en la universidad o en la profesión escogida.

Entre más variables entren en consideración, menos se concentra el trabajo escolar en una sola prueba y se tiende más a una educación de tipo general, pues la escuela dispone de un tiempo muy limitado para dedicarse a una determinada prueba o medición. Pensamos que es necesario desarrollar en los estudiantes algunas destrezas para tomar pruebas; sin embargo, se considera que la mayor parte del currículo escolar está más allá y por encima de cualquier prueba. En un estudio, actualmente en preparación, sobre el Programa de Matemáticas de Dallas, se encontraron en muchas escuelas materiales del currículo preparados por editores privados, que estaban diseñados para ayudar a las escuelas a pasar las pruebas estatales. Un grupo de profesores de cuarto curso no enseñó la materia siguiendo el libro de texto aprobado oficialmente porque opinó que no preparaba a los estudiantes para pasar las pruebas del estado. Consideraron que lo que no estaba incluido en las pruebas no debería enseñarse para evitar confusiones a los estudiantes (Bearden,1997). Este tipo de incongruencias son las que se deben evitar utilizando un amplio arreglo de variables en el sistema de valor agregado.

Después de la selección de variables, el componente más importante al que se debe prestar atención es el de la matrícula continua. Esto se refiere a los resultados que serán tenidos en cuenta en el sistema. En la evaluación de la escuela, un componente básico señalado en el Modelo CIPP ha sido la evaluación del proceso. Se considera inapropiado atribuir los resultados de un programa a su tratamiento si el programa no ha sido ejecutado. De manera similar, en los sistemas de eficacia de la escuela, si un estudiante se muda a una escuela un tiempo después de haber comenzado el año escolar, los efectos educativos asociados con este estudiante no pueden ser atribuidos a esta escuela. Es decir, la eficacia de las escuelas deberá ser

determinada por el desempeño de los estudiantes que estuvieron matriculados en la escuela la mayor parte del año escolar. Basar los resultados en estudiantes que se trasfirieron a la escuela unas semanas antes de las pruebas solamente desorienta. Es por tanto necesario establecer un período de tiempo mínimo durante el cual los estudiantes puedan ser admitidos en la escuela antes de que los resultados de ese estudiante puedan ser atribuidos a esa escuela en particular.

El quinto requisito es que los sistemas de evaluación de la escuela deben estar basados en cohortes de estudiantes. Los puntajes obtenidos por los distintos estudiantes a través del tiempo están sujetos a fluctuaciones que nada tienen que ver con la eficacia de la escuela. Además, los sistemas deben estar diseñados de tal manera que no puedan tomar ventaja de los altos o bajos puntajes que presentan los estudiantes en el momento de su ingreso. Este requisito implica la necesidad de utilizar metodologías estadísticas para predecir e interpretar los resultados de los estudiantes.

El otro componente del sistema de valor agregado es la norma sobre el porcentaje de participación en la prueba. Una escuela puede esconder algunos estudiantes cuando no existe una presión para evitarlo. Para las pruebas que requieran aplicarse en determinados períodos, debe existir una norma sobre el número específico de estudiantes elegibles para la prueba que deben tomarla. Nosotros, basados en nuestra experiencia de más de una década en Dallas (Dallas public Schools,1996), recomendamos un 95 por ciento. Para pruebas sin períodos concertados de antemano—por ejemplo la de Texas (TAAS)—se espera que las escuelas logren el promedio diario de asistencia menos un porcentaje. Los autores recomiendan este promedio menos un dos por ciento. Estas normas permiten asegurar que las escuelas no intenten engañar al sistema seleccionando ciertos estudiantes para las pruebas. Los autores han observado que es típico que las escuelas intenten incumplir esta norma, ocultando a los estudiantes con bajos resultados. En un sistema de valor agregado esta estrategia puede fallar ya que los estudiantes con bajos resultados también pueden agregar a la medida de eficacia de la escuela tanto como lo hacen los que tienen resultados altos. La estrategia de esconder a los estudiantes con bajos resultados funciona solamente cuando se emplea un sistema no ajustado o cuando la escuela proporciona servicios inadecuados a los estudiantes con bajos resultados. La norma del porcentaje de personas que deben participar en la prueba controla la situación anterior. Además, la información de la evaluación de valor agregado podrá ser explícitamente analizada considerando diversos grupos de estudiantes para determinar si todos están recibiendo atención suficiente.

También es esencial que el sistema de evaluacion sea imparcial. Esto significa que es necesario controlar por medio de metodologías adecuadas, el efecto de las variables referidas a los antecedentes de los estudiantes y sobre las cuales la escuela no ejerce un control directo. Estas variables incluyen aspectos de raza, género, dominio del lenguaje, nivel socioeconómico y cualquier otro aspecto que de alguna forma esté relacionado con los resultados que nos interesan. Es necesario también identificar las variables que influyen en los resultados a nivel de la escuela pero que no están

bajo su control. Entre las variables relacionadas con el tema de la imparcialidad se pueden mencionar las siguientes: movilidad estudiantil, condiciones de superpoblación, ingreso promedio de la familia, porcentaje de estudiantes de bajo nivel socioeconómico, porcentaje de estudiantes de diferentes razas, porcentaje de días de instrucción perdidos ya sea por motivos de salud o por ausencias del profesor.

TRES NIVELES DEL MODELO DE EVALUACIÓN DE LA ESCUELA

El modelo de evaluación de la escuela que se propone en este capítulo y que ha sido exitosamente ensayado en varios lugares, es un modelo a tres niveles. El primer nivel se relaciona con la escuela y ha sido diseñado para hacer a cada escuela responsable de la mayor parte de los aspectos de su funcionamiento. Una mayor autonomía escolar facilita la maximización de recursos, pero conlleva la posibilidad de que determinados grupos de personas en la comunidad escolar pueden verse perjudicadas. Por esta razón el concepto de responsabilidad debe ser el elemento fundamental de los planes para incrementar el proceso de toma de decisiones locales. Cada escuela debe recibir datos útiles para tomar sus decisiones, pero debe ser responsable de tales decisiones.

El segundo nivel se relaciona con el distrito y se ejecuta a través del Plan de Mejoramiento del Distrito (District Improvement Plan, DIP). El DIP establece los objetivos de instrucción y los niveles de rendimiento deseados, así como también señala la forma en que el organismo central va a apoyar a las escuelas. Dependiendo de la manera como el sistema educativo esté organizado en cada país, el DIP puede servir como un plan de desarrollo a niveles regional, estatal o nacional.

El tercer nivel del sistema incluye el mejoramiento escolar, o el uso de los índices de eficacia. Estos índices tienen en cuenta importantes variables sobre antecedentes del estudiante y ofrecen información acerca de cómo están trabajando las escuelas con sus estudiantes. Los componentes de los sistemas SIP (System Improvement Plan, Plan de mejoramiento del sistema) y DIP (District Improvement Plan) se preocupan de los resultados finales y los índices de eficacia de la escuela (School Effectiveness Indices, SEIs) añaden al sistema un componente de valor agregado.

Es muy importante enfatizar nuevamente que las dos características más importantes de la evaluación de una escuela o de la responsabilidad del sistema deberán ser la imparcialidad y utilidad. Los educadores que apoyen el movimiento de responsabilidad tienen derecho a que los estándares con que van a ser juzgados sean justos y objetivos. Es esencial que este sistema también ofrezca datos útiles para la toma de decisiones y el mejoramiento. El sistema esbozado en este capítulo incorpora el concepto de imparcialidad tal como está definido en el texto de los Estándares para la Evaluación de Programas (Comité Conjunto sobre Estándares para la Evaluación Educativa, Joint Committee, 1994) y en los Estándares para Pruebas Educativas Psicológicas (AERA, APA, NCME, 1985). También producen

datos para la toma de decisiones y el manejo a cada nivel del sistema.

En relación con el sistema para el establecimiento de responsabilidades es importante anotar que éste ha sido diseñado para aplicarse en el contexto de una administración local. Las escuelas públicas de Dallas están ejecutando un plan de educación centrada en la escuela, a través del Programa de Desarrollo Escolar del Centro de Estudios sobre la Infancia de Yale (Comer,1988). De acuerdo con este modelo, el director, los padres y los empleados participan en el proceso de toma de decisiones y en el gobierno de la escuela por medio del Consejo Escolar Comunitario (School Community Council, SCC), que es el encargado de tomar las decisiones apropiadas sobre el funcionamiento de la escuela. En cada escuela puede existir un gran número de comités, pero el SCC es el responsable del currículo, la enseñanza, la evaluación y otras medidas que afecten a todo el sistema. También se encarga del desarrollo de destrezas por parte de los padres y empleados, de la socialización, de la interacción escuela-comunidad y de las relaciones públicas. En el caso de la enseñanza secundaria estos comités incluyen también estudiantes.

Sin tener en cuenta la estructura de la respectiva organización, las funciones de evaluación que se llevan a cabo a nivel de la escuela incluyen: el desarrollo del SIP; la interpretación de los datos formativos para la solución de problemas y de los datos sumativos para reorientar prioridades, programas y recursos; la constancia sobre la ejecución de proyectos y programas dentro de la escuela, incluyendo el seguimiento a la ejecución del SIP; y la coordinación de toda la acción investigativa basada en la escuela . Los trabajadores de la oficina central de investigación serán los encargados a niveles distrital, regional, estatal o nacional de ofrecer capacitación al personal de la escuela para desempeñar las tareas antes mencionadas.

En otras palabras: el Plan Distrital de Educación Centrado en la Escuela se encarga del control de los recursos disponibles y de las decisiones relativas a la enseñanza a nivel de la escuela (Edwards,1991). Las únicas decisiones que los comités escolares no están autorizados a tomar, son aquellas que involucran la naturaleza y magnitud de los resultados por los cuales ellos son responsables. Un paso de suma importancia en el proceso de mejoramiento de la escuela consiste en determinar los indicadores de desempeño que permitan informar a los educadores, padres y miembros de la comunidad, sobre si los estudiantes están alcanzando o no un progreso satisfactorio en aspectos fundamentales del aprendizaje académico. El Comité de Trabajo para el Establecimiento de Responsabilidades, inspirado en el Sistema Estatal de Indicadores de Excelencia Académica, determina los indicadores de desempeño. Este sistema de indicadores es la base para la acreditación de las escuelas en el Estado de Texas. Los indicadores de responsabilidad son consistentes a través de los tres niveles de responsabilidad del sistema.

La clave para el éxito del sistema descrito es el Comité de Trabajo para el Establecimiento de Responsabilidades. Este Comité consta de veintisiete miembros propuestos por el Consejo de Educación, que es el encargado de vigilar el Sistema Distrital de Responsabilidad. Incluye entre sus miembros a cuatro profesores de

enseñanza primaria , siete de educación secundaria, cuatro directores de escuela, cuatro padres de familia, cinco miembros del sector empresarial y tres administradores de la oficina central. Además, cada una de las organizaciones de empleados tiene un miembro *ex officio* en el Comité de Trabajo. Este Comité se ocupa de varios de los aspectos de la responsabilidad del sistema incluyendo la metodología, las pruebas de rendimiento, la determinación y ponderación de las variables de rendimiento y la definición de las normas para la asignación de los premios en efectivo que están relacionados con este sistema. El Comité de Trabajo para el Establecimiento de Responsabilidades también recibe las quejas y preocupaciones relacionadas con la responsabilidad del sistema. La formación de un grupo de supervisión que se encargue de vigilar la responsabilidad del sistema es fundamental para lograr que éste sea justo y equitativo.

Primer nivel: proceso de mejoramiento de la escuela

El principal instrumento para el mejoramiento de la escuela debe ser el Plan de Mejoramiento de la Escuela (School Improvement Plan, SIP). Los SIPs están organizados alrededor de metas referidas a resultados que están directamente enfocados hacia las prioridades de la escuela. Las prioridades de ésta deben estar relacionadas directamente con las prioridades del distrito, región o estado. Cada escuela debe hacer lo que le corresponda para alcanzar los objetivos del distrito o del estado. Las metas del SIP pueden incluir los siguientes aspectos: (1) desempeño del estudiante en lenguaje (vocabulario, lectura, destrezas verbales y escritas); (2) desempeño del estudiante en matemáticas (solución de problemas, conceptos y destrezas en cómputo); (3) desempeño del estudiante en estudios sociales; (4) desempeño del estudiante en ciencias; (5) participación de los padres y la comunidad en las escuelas; (6) tasas de promoción; (7) estudiantes matriculados en cursos avanzados y en programas de honores; (8) tasas de graduación (tasas de deserción); (9) participación y desempeño de los estudiantes en las pruebas de ingreso a la universidad; (10) asistencia de los estudiantes; (11) asistencia de los profesores, y (12) clima escolar y seguridad.

Como parte del SIP, cada escuela debe desarrollar planes estratégicos de acción para cada meta. Cada plan de acción incluye los siguientes elementos:

- *Necesidad*. Un resumen de la evaluación de necesidades incluyendo una descripción del estado actual de la meta propuesta.
- *Fines*. Objetivos mínimos u otros estándares de desempeño que la escuela debe alcanzar para poder aplicar el plan. Tales objetivos están directamente relacionados con las metas del distrito o del estado.
- *Narración de la estrategia*. Resumen de lo que se propone hacer para alcanzar las metas.
- *Exenciones*. Una especificación de las exenciones en la política del distrito o del estado, para ejecutar la estrategia.
- *Actividades/Calendarios/Personal responsable*. Actividades, tiempos

correspondientes y responsabilidades del personal para lograr las metas de la escuela.

- *Seguimiento*. Metodología para dirigir, evaluar, ajustar y documentar actividades formativas para alcanzar las metas.
- *Recursos*. Un resumen de su distribución (por ejemplo, cambios requeridos en el personal para ejecutar las estrategias, dinero).

La figura 1 muestra un ejemplo de las metas de un SIP. Cada escuela recibe sus propios datos sobre cada una de las metas y es responsable de alcanzar los resultados propuestos. Las metas se refieren a criterios en el sentido de que las escuelas establecen sus propias metas en términos absolutos y aplican los recursos de que disponen para lograrlas. Los datos de la figura 1 están incluidos en el SIP. Pero existe una gama amplia de datos adicionales disponibles para que las escuelas puedan hacer el diagnóstico necesario sobre los aspectos más débiles de los estudiantes (en la figura 5 se encuentran algunos ejemplos de estos datos.)

Figura 1. Ejemplo de perfil de las escuelas de secundaria. Características de las metas SIP

Un problema frecuente de los sistemas que se inspiran en metas absolutas es la forma como esas metas se han establecido. En muchos casos estas metas expresan más un deseo que una posibilidad real. Los educadores se enfrentan con el dilema de proponer metas muy bajas y de ser acusados de plantear expectativas bajas, o de proponer metas muy elevadas que tienen muy pocas probabilidades de lograrse. El problema comienza a hacerse particularmente difícil cuando parte de la evaluación de un individuo se apoya en el grado de logro de esas metas. El componente de mejoramiento o índice de eficacia de este sistema, será descrito en la próxima sección de este capítulo y puede servir de punto de partida para establecer metas importantes basadas en las mejores prácticas encontradas. Esta metodología permite al sistema proponer logros que presenten un reto al personal de la escuela, pero que sean posibles de alcanzar.

La figura 2 muestra esquemáticamente el proceso de mejoramiento de la escuela tal como funciona dentro de los parámetros para la toma de decisiones locales. El proceso comienza en cada escuela con la evaluación anual de sus necesidades ya que se considera como un prerrequisito para el mejoramiento del conocimiento de los niveles de desempeño existentes. De ahí que un programa comprensivo de evaluación del contexto (evaluación de necesidades) sea la columna vertebral de cualquier sistema de renovación. La evaluación de contexto proporciona una línea de base de información que esboza el medio ambiente, describe las condiciones actuales y deseadas, identifica las necesidades insatisfechas y las oportunidades que no se han aprovechado, diagnostica los problemas que dificultan la satisfacción de necesidades y el aprovechamiento de las oportunidades. Un sistema adecuado de evaluación de contexto se apoya en una base de datos longitudinal que se mantiene a nivel del distrito o del estado y que ofrece información periódica sobre variables tales como la deserción escolar, la asistencia, los niveles de rendimiento, los patrones demográficos y vocacionales, el nivel socioeconómico comunitario, los patrones de valores dominantes y las características académicas del profesor. De este modo el sistema de evaluación de contexto ofrece las bases necesarias para formular cambios objetivos identificando las necesidades y, en algunos casos, proponiendo soluciones prácticas para las áreas conflictivas identificadas.

La figura 3 presenta un panorama general de los componentes básicos del contexto operacional de la evaluación de sistemas. Además de las investigaciones que se proponen en la figura 3 que deben realizarse anualmente, se pueden satisfacer también otras necesidades específicas de evaluación cuando se relacionan con las áreas que presentan problemas específicos. Como ejemplos pueden incluirse la evaluación del nivel de individualización de las experiencias de aprendizaje en la escuela, encuestas sobre el uso de drogas en la escuela, estudios sobre las percepciones de la comunidad educativa y los estudiantes acerca del valor y la eficacia de las prácticas actuales y los análisis de propuestas. Las escuelas también deberán tratar de suplir otras necesidades de información evaluativa con estudios locales y aplicación de pruebas y, en general, con el desarrollo de una agenda de evaluación.

Figura 2. Esquema del proceso de mejoramiento de la escuela

Figura 3. Ejemplo de estudios anuales de evaluación de contexto

3.1 *Perfiles de medición*. Un resumen de los resultados del programa de pruebas basado en normas aplicado en todo el sistema, además de datos referentes a las condiciones socioeconómicas de la comunidad y el perfil del personal docente. Los resultados deben agregarse a nivel de la comunidad escolar y presentarse teniendo en cuenta su relación con las normas nacionales. Estos perfiles son de naturaleza descriptiva y se utilizan para informar a los educadores y funcionarios sobre la calidad de la educación en el Distrito, la Región, o el Estado y diagnosticar de manera general las debilidades del programa docente.

3.2 P*erfiles de pruebas referidas a criterios*. Resumen de los resultados referidos a los criterios de los programas de pruebas. Se utilizan como complemento de los Perfiles de Medición y ofrecen estimados acerca de la alfabetización funcional de los niños en la escuela.

3.3 E*studios de seguimiento de los egresados*. Series de estudios sobre los que se graduaron en las escuelas. Incluye datos relacionados con las opciones de empleo de los que se graduaron, actitudes hacia la educación, nivel de vida, etc. Se usan para determinar qué programas satisfacen las necesidades de los estudiantes. Los datos pueden utilizarse como guía para la planeación del currículo. Deben hacerse seguimientos anuales y quinquenales.

3.4 *Estudios sobre deserción*. Series de estudios diseñados para ofrecer datos descriptivos relacionados con la deserción, variables asociadas, interacciones entre variables y tendencias de la deserción. Puede enfatizarse en aspectos relacionados con una temprana identificación de los desertores potenciales con el propósito de trabajar estrategias de intervención.

3.5 *Estudio de variables de insumo*. Seguimiento constante de los insumos de la escuela. Estos estudios ofrecen datos y análisis sobre costos, costo beneficio, costo eficacia, así como también información acerca del ambiente escolar.

3.6 *Estudios de medición*. Serie de estudios sobre la validez, confiabilidad y comparabilidad de las pruebas usadas. Estos estudios ofrecen información sobre el grado de confianza que se puede tener en los datos de la prueba.

3.7 *Estudios longitudinales*. Series de estudios encaminadas a determinar el rendimiento, la matrícula y las tendencias durante ciertos períodos de tiempo establecidos. Estos estudios ofrecen una medida de responsabilidad del sistema.

3.8 *Estudios de contexto de los estudiantes*. Series de estudios que buscan determinar los patrones de matrícula así como también la matrícula en programas de honores.

3.9 *Indices de eficacia de la escuela y los profesores*. Sistema diseñado para producir estadísticas ajustadas sobre los logros de los estudiantes en pruebas referidas a normas y criterios, así como también sobre otros aspectos importantes de la instrucción. El sistema identifica los profesores y las escuelas que tienen un mejor desempeño, así como también a los que tienen más bajos resultados. De este modo se prepara el terreno para futuros estudios. Los índices de valor agregado son tema de esta monografía.

Para obtener información completa acerca de las necesidades de decisión para la planificación, el sistema de evaluación de contexto debe tener la capacidad de

ofrecer proyecciones válidas sobre el futuro considerando ciertas variables importantes. La figura 4 señala áreas generales para estudios proyectivos que permiten obtener información crucial sobre tendencias futuras útiles para la mayor parte de las decisiones educativas del presente.

Figura 4. Ejemplos de estudios de proyección

4.1 *Estudios sobre aspectos demográficos referidos al estudiante y la matrícula*. Estudio diseñado para localizar y proyectar la población, proyectar la futura matrícula escolar, con el propósito de ofrecer la información necesaria para la planeación a largo plazo, para determinar tendencias y demandas posibles acerca del personal, los programas y las facilidades educativas.

4.2 *Estudios sobre el flujo de profesores*. Estudios diseñados para proyectar la cantidad y el costo de los profesores que se necesitan bajo una política de multiplicidad o de cambios ambientales. El estudio proyectará el número y características de los profesores que se quedan, que renuncian y que va a ser necesario contratar. Esta información es útil para evaluar las formas de contratación de los profesores, hacer proyecciones de personal y análisis de contratación, así como para proponer legislación evaluativa.

4.3 *Estudio sobre facilidades*. Estudios diseñados para proyectar la cantidad, el tipo y el costo de las áreas y espacios necesarios, establecer comparaciones con los existentes y determinar las deficiencias o excesos en cada escuela y sector demográfico. Esta información alimenta las decisiones sobre apoyo y construcción para las escuelas de la zona.

4.4 *Estudios financieros*. Estudios diseñados para obtener proyecciones financieras generales acerca de las necesidades del Distrito utilizando los insumos anteriores. Incluye aspectos tales como: proyección de las necesidades de ayuda y financiamiento, análisis de las deudas y requerimientos de nuevos bonos, análisis de la renta y los gastos, análisis de las tasas de impuestos.

Este tipo de estudios abarca muchas variables y están diseñados para ayudarnos a tomar decisiones inteligentes acerca del futuro, apoyados en información sólida. Además, para atender solicitudes particulares, pueden diseñarse otros modelos de proyección relacionados con problemas específicos, tales como el inventario y la reorganización de la cafetería; estos proyectos tienen una alta prioridad en la asignación de fondos.

Una vez que el sistema de evaluación de contexto ha identificado las necesidades, quienes toman las decisiones a nivel local deben priorizar las necesidades y concentrarse en ellas para reducir las discrepancias entre las condiciones deseadas y las existentes, señalando las metas para aquéllas necesidades consideradas altamente prioritarias. Es en este momento cuando la información de evaluación de insumo presta su mayor contribución. La evaluación de insumo es la que proporciona la información para determinar los métodos y la forma como se van a utilizar los

recursos necesarios para alcanzar las metas. En un sistema operacional existen seis recursos importantes de información acerca de los insumos:

- Información previa del producto de la evaluación sumativa, incluyendo los índices de eficacia de la escuela.
- Revisión de la literatura.
- Información de la investigación básica.
- Información de la investigación aplicada.
- Investigación-acción.
- Información adicional a la investigación o evaluación

La información proveniente de la evaluación sumativa de producto busca identificar hasta dónde las metas de un proyecto, programa o escuela han sido alcanzadas. Cuando existen evaluaciones de producto, referidas a un programa con metas similares a aquéllas identificadas en respuesta a la información de contexto, esta información suministra insumos útiles a quienes toman las decisiones para determinar la probabilidad de que el programa pueda reducir las discrepancias identificadas.

La información de la investigación básica se refiere a las relaciones fundamentales que afectan el aprendizaje del estudiante. Antes de tomar la decisión de ejecutar un programa, sus responsables deben juzgar hasta qué punto tal programa es coherente o no con los principios establecidos por la investigación básica sobre el aprendizaje y el desarrollo humano. Este proceso requiere una constante revisión de la literatura.

La información obtenida por la investigación aplicada se refiere a la interacción entre las características de los estudiantes, las del profesor y los sistemas de instrucción. La investigación aplicada se diferencia de la básica en que la información que arroja está más directamente relacionada con decisiones específicas en el contexto de la práctica educativa. Quienes toman las decisiones necesitan información relacionada con los tipos de estudiantes (clasificados, por ejemplo, por su nivel de ansiedad) que trabajan mejor en determinados sistemas de instrucción, manejados por profesores que a su vez poseen diversos tipos de características o razgos de personalidad.

El personal de las escuelas locales puede ser motivado y capacitado para diseñar, realizar e interpretar estudios de investigación-acción. El movimiento en favor de la autonomía escolar va a crear múltiples necesidades adicionales que no pueden ser satisfechas por la información producida centralmente. La investigación-acción como proceso de solución de problemas es muy apropiada para responder a necesidades locales y ayudar a construir su capacidad. Se trata de un proceso de selección de acciones y estudio de sus consecuencias para así poder tomar medidas más eficaces (Lewin,1946; Town,1973). La investigación-acción requiere de un continuo reciclaje a través de cuatro pasos: (1) identificación de necesidades, (2)

desarrollo de planes de acción enfocados a atender las necesidades, (3) ejecución de los planes de acción y (4) evaluación formativa y sumativa de los planes. En organizaciones abiertas como las escuelas, la fuerza de la investigación-acción descansa en su ejecución por parte de los miembros de la organización en sus respectivos lugares de trabajo. En efecto, los miembros de una organización aprenden activamente cuando estudian problemas en ambientes que ellos perciben como relevantes o importantes. Los resultados se utilizan para ofrecer la mayor cantidad posible de información formal a los departamentos de evaluación del distrito o del estado.

Finalmente, existe información que no proviene ni de la investigación ni de la evaluación, pero que debe formar parte de la mayoría de las decisiones educativas. Por ejemplo, información sobre la capacidad del personal, los costos, la viabilidad política para ejecutar un programa y las instalaciones existentes.

Una vez que la información de insumo se haya recolectado y preparado, se integra a la primera etapa de desarrollo del programa. Los responsables del programa determinan si hay suficientes recursos disponibles para hacer los cambios deseados en la instrucción. Con cierta frecuencia los recursos adecuados no están disponibles y es necesario hacer algunos compromisos. En muchos casos la carencia de recursos no está limitada sólo por problemas de costo y factibilidad política, sino también por una insuficiente base de información investigativa. Entonces los educadores se enfrentan a la dificultad de tener suficientes recursos materiales pero insuficientes recursos de información.

Si no se dispone de suficientes recursos materiales, el sistema puede tener que subsistir por un período de tiempo en un estado de "persistencia iluminada" por la información proveniente de la evaluación. Las evaluaciones de contexto periódicas continuarán arrojando luz sobre las diferencias entre lo que se desea y lo que existe. Si el problema proviene de una insuficiente cantidad de recursos de información, los programas se suelen ejecutar sin un apoyo suficiente de datos. La información de base se construye entonces, a través de una serie de estudios de investigación aplicada y evaluación sistemática.

Para superar el problema de una insuficiente información de recursos, se deberán establecer centros de desarrollo nacional, con la responsabilidad de desarrollar sistemas de instrucción que satisfagan las necesidades identificadas en la evaluación de contexto. Los materiales y sistemas de instrucción se desarrollan a nivel local únicamente cuando no existe otro material útil, puesto que el desarrollo de sistemas de instrucción es una tarea muy costosa.

La fase de planificación se inicia cuando se dispone de suficiente material y recursos de información, o cuando hay suficiente material y recursos mínimos de información. Esta fase se inicia como resultado de la recolección de la información realizada durante la evaluación de insumo. El evaluador al comienzo de esta fase de planificación del programa debe facilitar la mayor cantidad posible de la evaluación

de insumo a quienes planean el programa. Y una vez que se ha optado por un camino para remediar una necesidad existente, el evaluador debe asegurarse de que los objetivos propuestos para el programa sean mensurables.

Como complemento a las sesiones de planificación del programa, el evaluador desarrolla un plan detallado de evaluación, especificando los criterios a través de los cuales se va a juzgar la escuela. Todos los participantes e interesados en el proyecto deben tomar parte en el diseño. Este diseño empieza a constituir parte del Plan de mejoramiento de la escuela. El desarrollo del diseño de evaluación incluye necesariamente interacciones continuas entre los participantes en el proyecto y el evaluador con el fin último de incrementar al máximo la eficacia de la información. Obviamente, el evaluador debe mantenerse independiente de la administración del programa para asegurar una mayor objetividad en los resultados de su trabajo.

Una vez que la fase de ejecución está en marcha, el evaluador y el personal asignado para seguir la ejecución del SIP deben elaborar continuos informes de evaluación formativa relacionados con la aplicación del programa. Estos informes se refieren principalmente a dos categorías, evaluación del proceso y evaluación de productos intermedios. La evaluación de procesos tiene tres objetivos principales: (1) detectar y predecir las fallas del diseño de los procedimientos previstos o de su aplicación, durante la etapa de ejecución del programa; (2) proporcionar información para la toma de decisiones programadas; y (3) mantener un registro o memoria de los procedimientos realizados a lo largo del desarrollo del programa. Por medio de los datos provenientes de la evaluación de procesos, la gerencia de la escuela se mantiene informada sobre el grado de ejecución del programa de acuerdo con las especificaciones establecidas. Desde el punto de vista de la evaluación, esta información sirve de protección contra la formulación de juicios basados en acontecimientos ficticios.

La evaluación de productos intermedios ofrece a la administración de la escuela información periódica relacionada con el logro parcial de sus objetivos específicos durante la fase de ejecución. Tanto los informes de evaluación de procesos como los de evaluación de productos intermedios, informan a la administración del programa sobre las metas y los niveles logrados mientras aún es tiempo de hacer ajustes y cambios.

Una vez que se completa un ciclo en la ejecución del programa, se prepara una evaluación sumativa del producto. Este informe analiza hasta dónde se alcanzaron los objetivos relacionados a un conjunto específico de criterios especificados en el plan SIP y analiza también el costo-eficacia de los diferentes programas escolares realizados, en comparación con otras estrategias alternativas. La información relacionada con esta problemática se puede interpretar a la luz de la evaluación de productos intermedios. Sin información acerca de la ejecución de los programas, la información de la evaluación sumativa del producto es poco útil, puesto que sobre la base del producto solamente es muy difícil esbozar el curso de las actividades de mejoramiento de la instrucción.

Las escuelas deben promover el uso de carpetas, análisis de protocolos y otras formas de evaluación auténtica para seguir el desarrollo de sus programas. Esta información se puede usar para obtener evidencias sobre logros en instancias donde la mayor parte de los sistemas de evaluación comunes no son adecuados. Lograr este tipo de evaluación a nivel de todo el distrito o de todo el estado es difícil. Las pruebas de rendimiento se construyeron hace tiempo en el servicio de pruebas del distrito de Dallas (Assessment of Course Performance, ACP). Los ACPs son exámenes finales estandarizados para 72 cursos de los cursos 9 a 12. Durante una hora se resuelven pruebas de selección múltiple y durante otra se trabajan pruebas de desempeño. Estas pruebas fueron desarrolladas por el departamento de evaluación e incluían instrucciones detalladas para calificarlas. La porción relativa a las pruebas de desempeño era calificada por los profesores y verificada al azar por el departamento de evaluación. Las pruebas de desempeño han sido eliminadas por la administración del distrito debido a la gran cantidad de tiempo que consumen. Aunque no es seguro que se pueda lograr la confiabilidad necesaria para la calificación de dichas pruebas, es importante comunicar a los maestros el mensaje de que los tipos de destrezas y actividades que ellas miden son parte de las destrezas y actividades que el distrito desea que ellos enseñen a sus estudiantes. Desde esta perspectiva, las pruebas de desempeño son más un problema de currículo que de evaluación. Las primeras evidencias sobre las pruebas de desempeño señalan que son mucho más difíciles para los estudiantes que el promedio de las pruebas de selección múltiple (Dryden 1991). En la figura 5 se presentan listas de ejemplos de datos formativos y sumativos disponibles a nivel de las escuelas. Estos datos ofrecen mucha de la información de apoyo mencionada anteriormente con referencia a la figura 1.

La figura 5 presenta información sobre los indicadores, las posibles metas que el programa SIP se propone con cada indicador (las metas del SIP están contenidas en la figura 1, las audiencias importantes para cada indicador (1) estudiantes, (2) padres, (3) profesores, (4) director, (5) consejo escuela comunidad, (6) personal de línea de la oficina central, (7) personal de la oficina central, (8) Secretario de Educación y (9) Consejo de Educación y el público), el propósito principal (establecimiento de responsabilidades o toma de decisiones) y el momento de disponibilidad de la información.

El lector notará que la mayor parte de los datos que se incluyen en la figura 5 están relacionados con la toma de decisiones. Esto es consecuente con la filosofía de que la información sobre la responsabilidad del sistema es poco útil si no contribuye al diagnóstico y mejoramiento. En el diseño del sistema para el establecimiento de responsabilidades es importante analizar las necesidades de datos en cada punto de la organización, para enfocar los sistemas de información a la satisfacción de esas necesidades. Una forma de lograrlo—especialmente en lo relativo a los resultados—es identificar las necesidades de información por parte de los profesores y luego añadir otras, a medida que se consideran diferentes niveles en la escala de la organización, resumirlas y establecer las demandas de información de toda la

organización. Así se pueden satisfacer a tiempo las demandas de información, las de los profesores y las de los niveles más altos de la jerarquía del sistema.

Obviamente, si el personal de la escuela utiliza eficazmente los datos disponibles para el mejoramiento de la institución, se pueden logran propósitos fundamentales de capacitación. Primero, se debe enseñar al personal de la escuela a recoger e interpretar datos objetivamente. Segundo, se debe aprender a utilizar los datos disponibles en el diseño y la ejecución de los programas de instrucción. Se deben desarrollar módulos de capacitación para el personal de la escuela sobre aspectos relacionados con el diseño de carpetas y calificaciones en las pruebas de resultados, se debe hacer un análisis de protocolos, desarrollar pruebas hechas por los profesores, interpretar y utilizar los datos y diseñar y ejecutar proyectos de investigación-acción.

Segundo nivel: proceso de ejecución en el distrito

El segundo nivel de responsabilidad del sistema se concentra en el desarrollo del Plan a nivel del distrito (DIP), el cual presenta los objetivos y planes de acción estratégicos correspondientes, dentro de un horizonte de planificación a varios años. El plan satisface las necesidades básicas para el establecimiento de responsabilidades y los requerimientos de planificación estratégica de un gran número de personas interesadas incluyendo al Secretario de Educación, el Consejo de Educación, la Agencia de Educación del Estado, el personal de la escuela y el público en general. El DIP debe satisfacer los cuatro principales requisitos del sistema de planificación estratégica y recibir todos los insumos de todas las instituciones y departamentos del distrito. El DIP fija los objetivos para el establecimiento de responsabilidades de acuerdo a una serie de estándares de desempeño mínimos a nivel del distrito y de cada escuela y facilita a todo el sistema planes de acción que permitan alcanzar los principales objetivos del distrito; además, especifica la metodología requerida para alcanzarlos. Los objetivos DIP están en las mismas áreas que muchos de los objetivos SIP. Cada escuela debe cumplir con su parte para lograr los objetivos del distrito. Si el estado es el responsable de la educación, entonces el plan DIP deberá ser un Plan de desarrollo del estado.

El DIP contiene los planes estratégicos para ayudar a que las divisiones administrativas del distrito o del estado contribuyan al logro de los objetivos propuestos por el distrito o el estado. También incluye los niveles de resultados deseados en el último año previsto, así como los pasos intermedios necesarios para alcanzar los resultados deseados, a partir de la línea de base identificada previamente. El DIP está directamente relacionado con el SIP, puesto que los niveles de resultados especificados en el SIP deben ayudar al distrito a alcanzar sus metas. El DIP fija el nivel de los criterios para los resultados deseados. Las metas son absolutas, pero deben ser específicas para cada escuela. Todas las escuelas o ninguna de ellas pueden lograrlas, esto quiere decir que el rendimiento no se determina por una norma de grupo. Los objetivos DIP pueden ser establecidos también

empíricamente, apoyándose en las mejores prácticas.

Tercer nivel: índices de eficacia de la escuela

El último nivel para el establecimiento de responsabilidades del sistema es el más importante desde el punto de vista de la evaluación de las escuelas. Existen dos tareas particularmente difíciles en la evaluación de cualquier escuela: cómo definir la eficacia y cómo desarrollar un modelo para juzgarla.
En un intento de ofrecer una mejor definición de eficacia y de responder a la preocupación de muchos en torno a la limitada perspectiva de la investigación previa sobre esta última en la escuela, Murnane (1991), David (1987) y otros han señalado la importancia de desarrollar y ampliar varios de los indicadores de resultados. Además Oakes (1989), David (1987) y Cohen (1986) han discutido la importancia de incorporar los indicadores de insumo, proceso y contexto como aspectos importantes para desarrollar mejores mecanismos de responsabilidad.

Posibles indicadores de insumo serían: matrícula en la escuela, composición étnica y socioeconómica, proporción de niños con manejo limitado del lenguaje, matrícula en programas especiales, características del personal y recursos financieros. Los indicadores de proceso describen lo que se enseña y la forma como se enseña e incluyen: el consenso sobre los objetivos de la escuela, el liderazgo de instrucción, las oportunidades de aprendizaje, el clima escolar y el desarrollo del personal y la interacción entre los profesores. Los indicadores de resultado se refieren usualmente al efecto de la escuela en los estudiantes o a proporcionar información sobre otras definiciones de "buenas escuelas" y pueden incluir información acerca del rendimiento académico de los estudiantes, las tasas de asistencia de estudiantes y profesores, las tasas de deserción y terminación, el rendimiento de los estudiantes en el siguiente nivel de escolaridad, los niveles de satisfacción de padres y estudiantes, el porcentaje de estudiantes que completan cursos avanzados, la asistencia a la universidad y las metas individuales y de la escuela (David, 1987; Oakes, 1989; Olson y Webster,1986; Pollard,1987; Shavelson y otros, 1987).

Anatomía de los índices de eficacia

La metodología para juzgar la eficacia de la escuela propuesta en este documento, la define en términos de su desempeño excepcional—superior o inferior— en medidas de rendimiento, en relación con lo que se esperaría para todo su grupo de referencia (se trate del distrito, el estado o la nación). Cuando la población de estudiantes de una escuela se separa notablemente de las tendencias preestablecidas o de la tendencia general de estudiantes con similares características a las del grupo de referencia, este distanciamiento se atribuye al efecto de la escuela. El problema de medición del efecto de la escuela, se convierte entonces en establecer los niveles de logros alcanzados por los estudiantes en las variables más importantes de resultados y en determinar hasta dónde el promedio de estos estudiantes sobrepasa o está por debajo de las expectativas. Los procedimientos estadísticos empleados incluyen la

utilización de HLM para las variables a nivel de los estudiantes y el análisis de regresión múltiple para variables a nivel de la escuela, con el fin de calcular las ecuaciones de predicción a cada nivel y por cada variable de resultado (sin considerar la identificación de la escuela). Luego se utilizan estas ecuaciones con los estudiantes o las escuelas, para identificar las ganancias obtenidas de acuerdo a las expectativas. El Comité de Trabajo para el Establecimiento de Responsablilidades u otro grupo similar puede establecer ponderaciones a los diferentes resultados. Una vez que se han determinado los niveles ponderados de rendimiento, la metodología proporciona un indicador que señala cómo está rindiendo la escuela en relación a otras escuelas a través de su grupo de referencia. Hasta cierto punto los mismos objetivos seleccionados para los procesos SIP y DIP se utilizarán como variables de resultados en los índices de eficacia de la escuela. Así las escuelas trabajan para mejorar las variables que seleccionaron como metas absolutas por medio de sus planes SIP y son juzgadas en términos del logro de tales metas, de su mejoramiento y de su eficacia.

Una aproximación para desarrollar el modelo de evaluación de valor agregado en la escuela es el modelo de regresión de mínimos cuadrados (OLS, del inglés Ordinary Least Square). Este modelo de regresión se genera a partir de la ecuación estándar OLS (de mínimos cuadrados). Se representa como sigue en la ecuación 1 para las variables de nivel del estudiante:

Utilizando este modelo, la Y representa cualquiera de las variables de resultados en el sistema. La X representa la variable predictora, disponible para el modelo en cuestión. (Estos valores no se refieren a la escuela en el momento.) Después de que se despeja X, se aplica el modelo a cada estudiante y se determina el valor residual r_i . Este valor de r representa la porción del valor agregado del puntaje del estudiante, más cualquier error individual para ese estudiante sobre la medición de un resultado en particular Y. Esta ecuación se resuelve para cada posible variable Y y las diferencias entre los resultados obtenidos por residuos para cada uno de los estudiantes y cada una de las variables. A medida que se añaden predictores, incluyendo más variables antecedentes o concomitantes, se resuelven las ecuaciones de regresión múltiple estándar (OLS) en forma similar.

Una vez que se obtienen los residuos de los estudiantes, se determina la eficacia relativa para cada escuela a través de los siguientes pasos:

1. Se agrupan los valores de r_i para cada escuela.
2. Se suman los residuos y se determina la media de los residuos para cada escuela.

3. Se corrigen los residuos de la media para cualquier tamaño de muestra o para cualquier disminución que ajuste las medias por el tamaño de la muestra y la variación diferencial.
4. Se estandarizan las medias de los resultados individuales de Y en una escala que tenga una unidad común.
5. Se ponderan las medias para cada Y si las variables están ponderadas diferencialmente y están sumadas a través de las variables Y.
6. Se estandarizan y reajustan las medias ponderadas de acuerdo a una escala final, que ha sido escogida para el desarrollo de un puntaje de eficacia.

En forma abreviada, estos pasos representan la guía general del modelo de evaluación escolar basado en el valor agregado. Vale la pena resaltar que varios de estos pasos son indispensables en el proceso. El paso tres es crítico. Si las medias no se ajustan, ya sea por el tamaño de la muestra o por disminución, las medias resultantes tendrán diferentes varianzas. En este caso, las medias de las escuelas con tamaños pequeños de muestras estarán sesgadas, lejos de la media del distrito. Estas escuelas tendrán a priori la probabilidad de ser juzgadas más o menos eficaces que escuelas de mayor tamaño donde no se presentan otras diferencias.

El segundo paso crítico es el cuarto. Las medias de todas la variables deben estandarizarse con la misma media y desviación estándar antes de ponderarlas y sumarlas. Si esto no se hace, se puede estar poniendo demasiado énfasis en variables con varianzas muy grandes.

Una discusión más detallada del modelo de regresión OLS y sus aplicaciones puede encontrarse en Webster y otros (1994, 1995, 1996, 1997a). Estos documentos explican cuidadosamente el uso del modelo OLS tal como se aplica en el proceso para establecer responsabilidades en el sistema de valor agregado de las escuelas públicas de Dallas. El sistema de Dallas se discutirá en forma más detallada en este capítulo. Los artículos mencionados incluyen explicaciones más completas.

Existen muchas variaciones posibles dentro del marco general para un modelo de regresión OLS. No obstante, el lector debe estar consciente de qué significa cada variación y cómo puede afectar los resultados. Antes de hacer una descripción de los diferentes enfoques, vale la pena recordar que estos enfoques conllevan un mejoramiento significativo con respecto a la comparación entre escuelas cuyos resultados no han sido ajustados. Si un enfoque no controla todas las variables, quiere decir que es posible desarrollar un mejor enfoque de valor agregado. Cuando los autores señalan que una variable no está suficientemente controlada, se refieren por lo general a su comparación con otros modelos analíticos de valor agregado. Esto no significa que el nivel de control sea insuficiente en general. En muchos casos cuando dos valores son descritos como uno mejor y otro peor, ambos por lo general son aceptables y la diferencia es muy pequeña. Los tipos generales de modelos OLS son los siguientes:

1. El predictor está limitado a una medición anterior de la variable de resultado.

Esta es la situación más simple. Por ejemplo, el nivel de lectura del año anterior se utiliza como predictor del nivel de lectura del presente año y no se incluye ninguna otra variable. El problema con este modelo radica en que las variables concomitantes a nivel del estudiante no son controladas adecuadamente y un predictor de los resultados individuales de un estudiante está inadecuadamente controlado. Las variables concomitantes a nivel de la escuela, no son adecuadamente controladas y los predictores de resultados de un estudiante están sujetos a influencias anómalas, una de las cuales es las trampas que se hacen en los exámenes (Webster y otros, 1995). Desde un punto de vista positivo, un modelo con un sólo año de información previa conserva un gran número de los mismos sujetos en el análisis. Hemos encontrado que cada año que pasa arroja una pérdida del 8 al 10 por ciento de los datos. Este modelo también se puede usar para variables de resultado a nivel de la escuela. En este caso los pasos 1-3 son superfluos ya que se puede utilizar una regresión ponderada basada en el tamaño de la escuela.

2. Se usan dos o más años de variables antecedentes y no se incluyen variables concomitantes. Este es un modelo longitudinal. Las variables concomitantes referidas a la escuela y a los estudiantes no son adecuadamente controladas. Los datos que faltan se convierten en un problema (Webster y otros, 1997a). Por otra parte, el aspecto positivo es que las influencias anómalas son mejor controladas. Este modelo con predictores de dos años es el que recomendamos para variables de resultados a nivel de la escuela.

3. Se utilizan uno o más años de variables antecedentes a nivel de los estudiantes al igual que las variables concomitantes. Este modelo controla las correlaciones de los residuos con las características de los estudiantes. La correlación a nivel de las características de la escuela se ha mejorado, pero aún no se controla adecuadamente (Webster y otros, 1995).

4. Se usan uno o dos años de variables anteriores y variables concomitantes a nivel de los estudiantes y de la escuela. Este modelo controla bien las características de los estudiantes y ofrece mejor—aunque no absoluto—control a nivel de las variables de la escuela (Webster y otros, 1997a).

Al estudiar los diferentes modelos jerárquicos de regresión (HLM) de valor agregado, nuestra investigación se ha limitado al modelo de Bryk y Raudenbush (1992). Al no contar con versiones del modelo específicamente probados por otros investigadores en otras situaciones, limitaremos nuestras observaciones al enfoque jerárquico (HLM) que hemos probado y utilizado en nuestro modelo de valor agregado. Sin embargo, no vemos razones teóricas subyacentes por la cuales los resultados que hemos acumulado no puedan generalizarse a otras formulaciones con metodologías de modelos mixtos.

Las ecuaciones estándar para efectos al azar del modelo HLM se presentan en las ecuaciones 2-4 para un único predictor del nivel 1 y variables condicionantes para un predictor de nivel 2. Nótese que el nivel 1 contiene datos al nivel del modelo de

la escuela. Los dos tipos de datos son simultáneamente modelados en un modelo jerárquico (HLM).

La importancia de este punto se discutirá posteriormente durante la presentación del modelo. Como en el caso del modelo de regresión OLS, estas ecuaciones se pueden expandir por la inclusión de más variables predictores del estudiante a nivel 1 (X) y la inclusión de más variables condicionantes a nivel de la escuela, nivel 2 (W). Los efectos de la escuela se estiman directamente a partir de los residuos de Bayes empíricamente ajustados, resultantes de la aplicación del modelo HLM (Bryk y Raudenbaush, 1992). Nuevamente, nótese que nuestra investigación contiene formulaciones más explícitas del modelo bajo condiciones muy diferentes. El lector interesado podrá consultar a Webster y otros (1995), Mendro y otros (1995), Orsak y otros (1997), Weerasinghe y otros (1997) o Webster y otros (1996, 1997a), donde aparecen más detalles sobre los modelos y la discusión de sus aplicaciones.

Igual que sucede con el modelo OLS, existen también muchas posibilidades con el modelo HLM. Sin embargo, nosotros no daremos explicaciones más allá de la discusión relacionada con nuestra investigación en torno del modelo. Consideramos que el HLM ofrece un avance significativo en el control de las variables concomitantes a nivel de la escuela y su uso debe concentrarse en esta función. Generalmente, anotamos lo siguiente:

1. El modelo jerárquico HLM debe usarse como un modelo de efectos al azar, como se puede apreciar en las ecuaciones 2-4. Es posible, también, usar el modelo jerárquico HLM como un modelo de efectos fijos, lo que permite incluir una gran cantidad de variables a nivel de los estudiantes, pero puede producir una supresión de los efectos entre las variables. (Webster y otros, 1997a).

2. El uso del modelo jerárquico HLM como modelo al azar es el mejor enfoque para controlar variables específicas a nivel de la escuela y controlar también variables no específicas a nivel de los estudiantes (Webster y otros, 1997a). Esto

requiere la utilización de un modelo modificado para incluir muchas variables a nivel de los estudiantes, pues el modelo jerarquico HLM al azar no permite utilizar directamente un gran número de estas variables (Webster y otros, 1996, 1997a).

3. El uso del modelo jerárquico HLM sin variables condicionantes a nivel de la escuela, nivel 2, no ofrece beneficios significativos sobre el modelo OLS de valor agregado en relación con la eficacia de la escuela (Mendro y otros, 1995; Weerasinghe y otros 1997; Webster y otros, 1995, 1996, 1997a).

4. No hay diferencia significativa entre el modelo jerárquico HLM basado en predecir un resultado apoyado en una variable predictora y los modelos basados en los puntajes ganados cuando se usa el pretest como predictor (Weerasinghe y otros, 1997). Sin un pretest como predictor, los modelos basados en puntajes ganados están correlacionados negativa y significativamente con el pretest.

MODELO DE VALOR AGREGADO DE DALLAS

En la figura 6 se pueden observar las variables utilizadas para calcular la ecuación sobre la eficacia de la escuela que se usa en las escuelas públicas de Dallas. Cada variable de resultados se describe bajo el título " resultados" junto con los cursos en los que se incluye, el puntaje resultado que es la base para el análisis, la metodología utilizada, el nivel al que se analizan los datos (nivel de la escuela o de los estudiantes), los predictores posibles, los cursos y las posibles variables condicionantes a nivel de la escuela que se incluyen en la ecuación a nivel de los estudiantes. Se utilizan dos modelos de regresión diferentes de acuerdo a la unidad de análisis que se vaya a utilizar. Si la unidad de análisis es el estudiante, se emplea el modelo lineal jerárquico; si la unidad de análisis es la escuela se utiliza el modelo de regresión múltiple. Por medio de estos enfoques es posible obtener predicciones altamente confiables acerca de los resultados de los estudiantes y de la escuela y hacer comparaciones entre los resultados actuales y los predichos. Todos los análisis que se hacen a nivel de los estudiantes se calculan sobre residuos, esto es, las estadísticas de las características de los estudiantes sobre las que la escuela no tiene control se quitan de la ecuación (género, raza, manejo del lenguaje, nivel socioeconómico e interacciones entre estas variables).

Si se usa el modelo jerárquico HLM como un modelo de efectos al azar, es necesario hacerle algunas modificaciones con el propósito de incluir un amplio grupo de variables a nivel de los estudiantes, como se anotó anteriormente. En el modelo de valor agregado de Dallas se hicieron estas modificaciones por razones tanto matemáticas como políticas (Webster y Mendro, 1995). La idea de incluir un grupo amplio de variables concomitantes a nivel de los estudiantes es la de eliminar los efectos específicos debidos a esas diferencias. Siempre hay escépticos entre los educadores y entre el público que creen que esto no es posible. Como respuesta parcial a este escepticismo en las primeras etapas del desarrollo del modelo de valor agregado de Dallas, nosotros adoptamos un sistema de regresión en dos etapas.

En este sistema las variables concomitantes a nivel del estudiante descritas en la figura 6 son regresadas tanto contra las variables de resultados como contra las anteriores medidas predictoras. Los residuos de estas regresiones se usan como efectos al azar en el sistema jerárquico HLM con un grupo de las variables condicionantes a nivel de la escuela. Desde el punto de vista matemático, sin este proceso de dos etapas, se requeriría un modelo jerárquico HLM de efectos fijos. Este no es el tipo de modelo que nosotros preferimos por las razones ya expuestas, específicamente por la posibilidad de efectos supresores. Desde el punto de vista político se pueden calcular los promedios de los residuos para las variables predictoras y las de resultados por categorías tales como: programas de nutrición, género, raza y manejo del lenguaje y puede demostrarse que son iguales. Pero aunque el escepticismo no se puede eliminar totalmente, se puede disminuir al demostrar que es posible controlar el efecto de estas variables (Webster y otros, 1997a).

Un segundo procedimiento específico del modelo de Dallas, es el de ajustar los residuos en la primera de las dos etapas. Después de que se calculan los residuos en la primera etapa, el espacio predictor se divide en 256 intervalos iguales y los residuos dentro de cada intervalo se estandarizan con una media de cero y una desviación estándar de uno. Este proceso asegura que, al cambiar la varianza de los residuos y las medias a través del espacio predictor, los estimados de valor agregado no se ven afectados (Mendro y otros, 1995; Weerasinghe y otros, 1997; Webster y otros, 1995, 1996, 1997a).

Se puede resumir el modelo de responsabilidad de valor agregado de Dallas en los siguientes puntos:

1. Se predicen las variables de la escuela por medio de una regresión regular OLS para la que se usan los datos de los dos años anteriores. Los puntajes sobre eficacia se calculan a partir de los residuos de la regresión.

2. Las variables de los estudiantes se predicen con un modelo de regresión OLS de dos etapas modificado y con el modelo de regresión HLM.

3. La primera etapa del proceso de análisis a nivel de los estudiantes regresa las variables de resultados y las variables predictoras anteriores contra las variables concomitantes a nivel de los estudiantes, ajusta los residuos por homogeneidad y proporciona los residuos para la etapa de HLM.

4. La segunda etapa del proceso de las variables de los estudiantes usa los residuos de un año anterior a la primera etapa para predecir los residuos de los resultados de la primera etapa en un modelo de efectos al azar HLM de dos niveles, incluyendo un grupo de variables condicionantes del nivel de la escuela, en un segundo nivel.

5. Se estandarizan y ponderan los resultados de cada análisis HLM por cada variable de resultados de los estudiantes y los resultados, a nivel de la escuela, de la

regresión OLS. Estas ponderaciones están basadas en las recomendaciones del Comité de Trabajo para el Establecimiento de Responsabilidades.

6. Los resultados ponderados se combinan para dar una estimación total de la eficacia de la escuela para cada escuela.

Algunos de los beneficios adicionales del modelo de Dallas (muchos de los cuales podrían aumentarse con la experiencia de otros modelos utilizados, según la estructura discutida anteriormente) son:

1. El sistema proporciona una clasificación de las escuelas de acuerdo a su rango en cada variable de resultados. Esto permite construir un perfil de la escuela, variable por variable.

2. Se pueden aislar y reagrupar los residuos de las variables de los estudiantes en el análisis HLM por categorías tales como: vinculación a programas de nutrición, género, raza y manejo del lenguaje, lo que permite a las escuelas determinar si los esfuerzos educativos están sesgados hacia un grupo particular de estudiantes.

3. Los requisitos del porcentaje mínimo de participación estudiantil en las pruebas ha producido resultados positivos. Hay un resultado en un alto nivel de asistencia a todas las pruebas y mediciones de resultados.

4. El uso de variadas mediciones de resultados ha reducido la tendencia de las escuelas a concentrarse solamente en una prueba. (Esto no significa que se haya eliminado la tendencia de las escuelas a concentrarse en las pruebas de acreditación del estado, pero la información disponible permite identificar las escuelas que se han concentrado mucho en una variable en particular).

USOS DEL SISTEMA DE VALOR AGREGADO EN LA EVALUACIÓN DE LAS ESCUELAS

Los resultados del sistema de valor agregado para establecer responsabilidades tiene muchos usos en la evaluación de las escuelas. Esto es posible gracias a las características del sistema, según las cuales los datos de los estudiantes se analizan a nivel individual, a la vez que se obtienen residuos por estudiante. De este modo es posible desarrollar estimaciones de eficacia de valor agregado para cualquier agrupación de estudiantes, por ejemplo: una clase, un programa o un proyecto.

Algunas veces esas agrupaciones disponibles se encuentran convenientemente en la escuela. Por ejemplo, un programa experimental de matemáticas que incluye a todos los estudiantes de cuarto curso de tres escuelas permite establecer una comparación de las estimaciones de valor agregado de los resultados en matemáticas para las tres escuelas. Dada la naturaleza de estas estimaciones, el distrito en su totalidad constituye un punto de referencia muy útil para la comparación de resultados específicos. Si las tres escuelas mencionadas en el ejemplo, están por

encima del distrito en cuanto a eficacia, se puede inferir que hay una evidencia de que el programa está teniendo un efecto. Combinadas con una evaluación de proceso y otras mediciones no incluidas en las mediciones de resultados (por ejemplo, pruebas específicas referidas a criterios), las mediciones de valor agregado contribuyen notablemente a la evaluación. En particular, los esfuerzos que incluyen a toda la escuela se pueden medir fácilmente usando los resultados de valor agregado para toda la escuela. En muchos casos, tales como la reestructuración de una o varias escuelas, la selección de escuelas para participar en programas de escuela-comunidad, o cualquier otro esfuerzo que pretenda afectar los resultados de la escolarización, las mediciones de valor agregado pueden proporcionar una potente dimensión a la evaluación.

Cuando se tienen grupos disponibles por programa, la recolección de la información por variables de resultado y programa sirve como referencia rápida. Por ejemplo, un programa en el que participan determinados estudiantes de una escuela puede ser fácilmente evaluado, tomando los residuos de los estudiantes que están en el programa y los de aquéllos que no están en el programa y comparándolos entre sí. Vale la pena resaltar que todos los resultados se refieren al distrito como un todo y que las estimaciones de valor agregado para los estudiantes que están o no en el programa ofrecen una idea imparcial acerca del mérito del programa.

Al utilizar un modelo similar al descrito anteriormente se eliminan de la ecuación la influencia de variables contextuales a nivel de los estudiantes y de la escuela, sobre las que la escuela no tiene control. Las escuelas no obtienen una ventaja particular al empezar con estudiantes blancos o con estudiantes de grupos minoritarios, o con estudiantes ricos o pobres, o con estudiantes con buen o pobre desempeño en el idioma, o con hombres o mujeres. Las escuelas que tienen grandes concentraciones de varias combinaciones de estudiantes tampoco están en ventaja o desventaja. Otras variables tales como la movilidad y la superpoblación, también son controladas a nivel de la escuela. La investigación sobre estos modelos ha mostrado que los resultados obtenidos no se correlacionan significativamente con ninguna de las variables contextuales a nivel de los estudiantes o a nivel de la escuela (Webster y otros, 1995, 1996,1997a). Por tanto, el campo no está limitado de antemano. Las preocupaciones de los investigadores sobre el impacto de las variables relacionadas con antecedentes sobre los efectos de la escuela se reducen notablemente desde esta perspectiva.

Además, dada la naturaleza de las ecuaciones de valor agregado, las escuelas no derivan ventajas particulares incluyendo estudiantes con altos o bajos puntajes. A los estudiantes con altos puntajes se les predicen altos puntajes. Para lograr esto las ecuaciones se deben desarrollar a nivel individual de los estudiantes.

A estas alturas, el lector se preguntará: la precisión adicional obtenida con el modelo de regresión basado en el modelo de valor agregado, ¿justifica las dificultades adicionales? ¿Por qué no se evalúan las escuelas con pruebas absolutas o

con los resultados no ajustados que se han obtenido? El siguiente estudio de caso, que los utilizó, ilustra las dificultades que surgen y los problemas de parcialidad inherentes a tales sistemas.

Una importante institución educativa, como muchas otras oraganizaciones de educación en los Estados Unidos, desarrolló su propio sistema para establecer responsabilidades. Este sistema estaba muy limitado a los resultados de las pruebas del estado referidas a criterios, asistencia de los estudiantes y tasas de deserción. El sistema diseminaba los datos sobre los distritos escolares de manera transversal y transversal-longitudinal y pretendía hacer comparaciones entre un distrito y otro y entre un distrito y el estado considerado como un todo. No utilizaba cohortes de estudiantes. Aunque no se pretende criticar este sistema, en él se encuentran excelentes ejemplos de metodología e interpretación inapropiadas.

La medida primaria sobre la cual se basa el sistema es las pruebas referidas a criterios. Han surgido serios reparos sobre la manera como se ha planteado la confiabilidad, la validez y las escalas de tales pruebas para los propósitos que se están usando. También con frecuencia se manifiestan preocupaciones relacionadas con el hecho de basar todo el sistema de evaluación en una prueba. La insistencia de no distribuir la información sobre las pruebas encubre al programa de pruebas bajo un velo de confidencialidad y secreto que sólo añade dificultad a la interpretación de los resultados e inconformidad en los usuarios. (Ejemplos de las pruebas se han empezado a entregar ahora después de una confrontación judicial.)

Aun pasando por alto los posibles defectos de la prueba, persisten dificultades acerca del establecimiento de responsabilidades del sistema. La primera se refiere a la arbitrariedad de las metas, es decir, las metas no fueron empíricamente establecidas y no tienen en cuenta los diferentes niveles de dificultad o características de las pruebas. La segunda es que la primera fase del sistema está basada sobre resultados de pruebas sin ajustar. Esto quiere decir que los estudiantes—y no las escuelas—están siendo evaluados. Como se mencionó anteriormente, la técnica de comparar escuelas sobre la base de resultados sin ajustar afecta adversamente a la medición de las escuelas con estudiantes que tienen características demográficas que difieren de la norma. Esto es especialmente cierto cuando se refiere a escuelas con grupos minoritarios numerosos y con poblaciones estudiantiles pobres.

Para ilustrar mejor este punto, más adelante se incluye el cuadro 2 en el que se muestran algunas de las características demográficas del veinte por ciento más alto de las escuelas, identificadas de acuerdo con la metodología descrita en este documento. A través de una comparación con otras escuelas del estado el lector puede observar que las escuelas eficaces, como se han definido con esta metodología, vienen en todas las formas y tamaños. También se incluyen estadísticas referidas al distrito para proporcionar un marco para la interpretación de los datos.

Cuadro 4. Características demográficas del veinte por ciento de las escuelas más eficaces

A nivel de escuela primaria (preescolar a 6 cursos) en las escuelas más eficaces el nivel de matrícula tiende a estar por debajo del promedio de matrícula de las escuelas de primaria del distrito. El tamaño de la matrícula oscila entre 193 para las más bajas y 860 para las más altas. Los grupos étnicos están representados en un 99 por ciento de negros, 90,4 por ciento de hispanos y 64,5 por ciento de blancos en los casos más altos; y 3,5 por ciento de negros, 0,3 por ciento de hispanos y 0 por ciento de blancos en los casos más bajos. La mayor parte de los índices de privación están por encima del promedio del distrito que es de 69, llegando hasta 92; mientras el porcentaje de estudiantes con idioma deficiente oscila entre un 57,9 por ciento para los casos más altos y 0 por ciento para los más bajos. En síntesis, no se puede predecir a partir de las características demográficas de los estudiantes que asisten, si la escuela estará o no clasificada entre las más eficaces. Esto se debe a que el sistema está basado en el mejoramiento y no en los niveles absolutos de logro y sus variables asociadas.

La última columna del cuadro 2 (SR) representa la posición de la escuela basada en el porcentaje de estudiantes que pasan todos los subtests de la prueba del estado referida a criterios. Las 27 escuelas primarias en el distrito consideradas más eficaces están entre los rangos 3 y 107, cuando esta jerarquización se basa en niveles de rendimiento absoluto. Debe notarse que las seis escuelas del distrito que quedaron ubicadas entre las quince mejores del sistema estatal, cuando se tomaron en cuenta fuentes de variacion desconocidas, tenían al menos el 50 por ciento de población estudiantil blanca y no tenían índices de privación por encima de 40. A nivel del estado, los más altos puntajes en este sistema son generalmente obtenidos por escuelas a las que atienden un alto porcentaje de estudiantes blancos y económicamente solventes.

En los cursos séptimo y octavo las escuelas más eficaces tienen una matrícula que oscila entre 367 y 888 donde, entre el 7 y el 30.3 por ciento son blancos, entre el 16,3 y el 75,4 por ciento son negros y entre el 15,7 y el 63,2 por ciento son hispanos. Los índices de privación oscilan entre un 30 y un 85 por ciento, el uso limitado del idioma varía entre 0 y 31,2 por ciento. Los rangos de las cinco mejores escuelas secundarias varían entre 2 y 10 en el TAAS. Estos rangos en cierta forma son muy cercanos, puesto que las características demográficas de las escuelas secundarias intermedias de Dallas son muy parecidas a las de las escuelas primarias; además, no existen muchas de estas escuelas.

A nivel de la escuela secundaria, las escuelas que tienen una organización especial llamadas "Magnet Schools" se destacaron. Cuatro de las seis escuelas consideradas como mejores participan en este programa. Este resultado era predecible ya que los estudiantes de estas escuelas (Magnet) tienden a estar muy motivados, puesto que tuvieron la oportunidad de escoger su escuela. La escuela secundaria considerada como la máseficaz en el distrito tenía un 97,9 por ciento de población negra y un índice de privación de 32 y fué ubicada en el puesto dieciséis entre veintiséis escuelas secundarias del Estado.

Los índices de eficacia de la escuela hacen referencia al tema y la práctica de la equidad y la imparcialidad. Las escuelas no tienen control sobre los estudiantes que reciben, pero sí tienen la responsabilidad de educarlos eficazmente. Las escuelas eficaces son escuelas buenas. Escuelas buenas son aquellas que mejoran con eficacia el rendimiento del estudiante en las mediciones que se consideran importantes.

UTILIZACIÓN DE LA INFORMACIÓN DE VALOR AGREGADO

El modelo de valor agregado para establecer responsabilidades de Dallas se ha venido utilizando desde 1991-92 para ofrecer mediciones no sesgadas acerca de la eficacia de la escuela. A partir de este principio se ha calculado y se ha entregado a las escuelas información más detallada sobre el sistema. La información que actualmente se ofrece incluye:

- Una medida totalmente ponderada acerca de la eficacia, lo que ofrece a la escuela, a la administración, a las directivas, y al público una respuesta sobre la pregunta "¿Cuál es la eficacia promedio de esta escuela cuando se tienen en cuenta todos los resultados considerados importantes?"
- Mediciones por grado o curso en relación a la variable de resultados sobre la eficacia de cada escuela en particular con:
  - Una diversidad de medidas por curso para ayudar a determinar la eficacia de cada equipo que trabaja con cada curso.
  - Una variedad de medidas agrupadas por categorías de materias importantes (lenguaje, artes, matemáticas, ciencias y estudios sociales) para ayudar a determinar la eficacia del núcleo de cada currículo.
  - Representaciones gráficas de la eficacia por cursos para brindar a los administradores que no están interesados en los números, un resumen eficiente de los resultados.
- Distribuciones de la eficacia de la escuela por curso para cada variable de resultado, desagregada por raza, manejo del lenguaje, género y nivel socioeconómico, con el propósito de mostrar si los grupos han sido justamente atendidos en la escuela.

Además, la División de Investigación, Evaluación e Información de Sistemas del distrito ofrece entrenamiento regular a los administradores, al consejo directivo, a las escuelas, a los grupos de padres, al Comité de Trabajo para el Establecimiento de Responsabilidades y a otras organizaciones sobre los índices, el uso general y específico de los resultados para la evaluación de las escuelas y el uso de los resultados para la planificación general o específica (Barden, 1997).

Los sistemas de evaluacion del distrito, de la escuela y del personal se han coordinado de tal manera que los resultados se pueden utilizar en la evaluación de los maestros (Bembry, 1997), en la evaluación del director y de los administradores escolares y en el desarrollo del sistema para el mejoramiento de la escuela (Webster, 1997b). En cada instancia se utiliza el modelo de evaluación en tres etapas. En la primera etapa se estudian todos los recursos de información y se prepara un análisis

de necesidades, con la condición de que se incluyan los resultados de valor agregado disponibles. En la segunda etapa se planean las estrategias, se documentan bien, y se especifica quién es el responsable de ejecutarlas, ya sea el profesor, el director o la escuela. Durante el año escolar estas estrategias se ejecutan, supervisan, evalúan y revisan adecuadamente. La tercera etapa corresponde a la evaluación del cumplimiento de las estrategias que utilizan la documentación que se especificó en el paso dos y que se recogió durante el año escolar. La ventaja de tener estos sistemas convenientemente alineados es que todos los datos se manejan con las medidas de valor agregado, conformándose así la fuente de información más importante (pero no única); el sistema trabaja enfocado hacia la razón de ser de la escuela: ayudar a los estudiantes a aprender. Todas las evaluaciones ya sean a nivel individual o escolar, así como también los planes, buscan mejorar la educación.

La implantación del modelo de los índices de eficacia escolar ha coincidido con un período de incremento general de los logros, reducción de las tasas de deserción, aumento de la asistencia y mejoramiento de otras medidas directamente relacionadas con el establecimiento de responsabilidad del sistema. Los índices forman la base de un programa de seis años de recompensas monetarias y reconocimientos a las escuelas eficaces y a su personal. Los índices son empleados regularmente por las directivas como una medida primaria para retener y seleccionar a los directores. Los índices se han usado en estudios sobre la eficacia o ineficacia de las escuelas y sus diferencias. Los índices han constituido el instrumento más importante para reestructurar las escuelas (Webster, 1997b). Los estudios sobre eficacia de la escuela que han utilizado los índices, han obtenido una idea real de la eficacia de la escuela, de su medio ambiente y del papel que juegan los directores y profesores en los aspectos que influyen en el grado de eficacia. Finalmente, como se aprecia en este capítulo, la información acerca de las escuelas y los estudiantes proporcionada por los índices de eficacia de la escuela se utiliza regularmente en los esfuerzos de evaluación de los programas del Distrito.

REFERENCIAS

Aiken, L. S. y West, S. G. (1991). *Multiple Regression: Testing and Interpreting Interactions*. Newburg Park: Sage.

Bano, Sm. M. (1985). *The Logic of Teacher Incentives*. Washington, D C: National Association of State Boards of Education.

Bearden, D. K. (1997). *An Overview of the Elementary Mathematics Program Evaluation Report REIS97-116-2*. Dallas: Dallas Public Schools (en progreso).

Bembry, K. (1997). *Implementation of the Dallas Public Schools Teacher Evaluation System: 1996-97* Report REIS97-120-2. Dallas, TX: Dallas Public Schools.

Bock, R. D. (1989). *Multilevel Analysis of Educational Data*. San Diego: Academic Press.

Bryk, A. S. & Raudenbush, S. W., Seltzer, M., y Congdon, R. (1988b). *An Introduction to HLM: computer Program User's Guide* (2nd ed.) Chicago, IL: University

of Chicago.

Bryk, A. S. y Raudenbush, S. W. (1992). *Hierarchical Linear Model: Applications and Data Analysis Methods*. Newberry Park, CA: Sage.

Bryk, A. S., Raudenbush, S. W., Seltzer, M., y Congdon, R. (1988a). *Toward A More Appropriate Conceptualization of Research on School Effects: A Three-level Hierarchical Linear Model*. Multilevel Analysis of Educational Data, San Diego: Academic Press.

Campbell, D. T., y Stanley, J. C. (1963) Experimental and quasi-experimental designs for research on teaching. En: N. L. Gage (Ed.), *Handbook of research on training*. Chicago: Rand McNally.

Cohen, J. (1968). *Multiple Regression As A General Data-Analytic System*. Psychological Bulletin, 70: 426-443.

Cohen, J. y Cohen, P. (1983). *Applied Multiple Regression/Correlation Analysis for the Behavioral* Sciences (2nd ed.) Hillsdale, New Jersey: Laurence Erlbaum

Cohen, M. (1986). Designing State Education Assessment Systems. *Study Group on the National Study of Student Achievement*.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McParland, J., Mood, A. M., Weinfeld, F. D., y York, R. L. (1966) *Equality of educational opportunity*. Washington, DC: U.S. Department of Health, Education, and Welfare, Office of Education, 1966.

Comer, J. P. (1988). Educating Poor Minority Children. *Scientific American*, 259, 5: 42-48.

Cook, D. L. Program evaluation and review techniques, applications in education. (1966) *U. S. Office of Education Cooperative Monograph*, 17, (OE-12024).

Council of Chief State School Officers (1995). *State Education Accountability Reports and Indicator Reports: Status of Reports Across the States*. Washington, DC.

Cronbach, L. J. (1963) Course improvement through evaluation. *Teachers College Record*, 64: 672-683.

Dallas Public Schools (1996). *School Performance Improvement Awards, 1996-97*. Dallas, Texas: DISD

Darlington, R. B. (1990) *Regression and Linear Models*. Nueva York: McGraw-Hill

David, J. (1987). Improving Education with Locally Developed Indicators. *Center for Policy Research in Education*. New Brunswick: Rutgers University.

Dempster, A. P., Rubin, D. B., y Tsutakawa, R. V. (1981). *Estimation in Covariance Components Models*. Journal of the American Statistical Association, 76: 341-353.

Dryden, M. E. (1991). *Evaluation of the 1990-91 South and West Dallas Learning Centers*. Dallas, TX: Division of Evaluation and Planning Services, DISD.

Edwards, M. E. (1991). *School-centered Education: A Plan for Creating a System of Schools in the DISD*. Dallas, TX: DISD

Eisner, E. (1976). Educational Connoisseurship and Criticism: Their Form and Functions in Educational Evaluation*", Journal of Aesthetic Education*, 3-4, 10: 135-150.

Elston, R. C., y Grizzle, J. E. (1962). Estimation of Time Response Curves and Their Confidence Bands. *Biometrics*, 18: 148-159.

Felter, M., y Carlson, D. (1985). Identification of Exemplary Schools on a Large Scale. En: Austin y Gerber (eds.), *Research on Exemplary Schools*. Nueva York:

Academic Press: 83-96.

Gallegos, A. (1994). "Meta-Evaluation of School Evaluation Models", *Studies in Educational Evaluation*, 20: 41-54.

Glass, G. V. (1978). Standards and Criteria. *Journal of Educational Measurement,* 15: 237-261.

Goldstein, H. (1987). *Multilevel Models in Educational and Social Research*. Nueva York: Oxford University Press.

Henderson, C. R. (1984). *Applications of Linear Models in Animal Breeding*. Guelph, Canada: University of Guelph.

Jaeger, R. M. (1992). Weak Measurement Serving Presumptive Policy. *Kappan*, 74, 2: 118-128.

Joint Committee on Standards for Educational Evaluation (1994). The Program Evaluation Standards. Western Michigan University, The Evaluation Center.

Kirst, M. (1986). New Directions for State Education Data Systems. *Education and Urban Society*, 18, 2: 343-357.

Klitgaard, R. E., y Hall, G. R. (1973). *A Statistical Search for Unusually Effective Schools*. Santa Monica, CA: Rand Corporation.

Laird, N. M., y Ware, H. (1982). *Random-Effects Models for Longitudinal Data*, Biometrics, 38: 963-974.

Lessinger, L. M. (1970). *Every kid a winner: Accountability in education*. Nueva York: Simon and Schuster.

Lewin, K. (1946). Action Research and Minority Problems. (*Journal of Social Issues*. 2: 58-73.

Lindquist, E. F. (Ed.). (1951). *Educational measurement*. Washington, DC: American Council on Education.

Longford, N. T. (1987). *A Fast Scoring Algorithm for Maximum Likelihood Estimation In Unbalanced Mixed Models with Nested Random Effects*. Biometrika, 74 (4): 817-827.

MacKenzie, D. (1983). School Effectiveness Research: A Synthesis and Assessment. En: P. Duttweiler (ed.), *Educational Productivity and School Effectiveness*. Austin, TX: Southwest Educational Development Laboratory.

Madaus, George F., Scriven, M. S., y Stufflebeam, D. L. (1983). *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*. Hingham, MA: Kluwer Academic Publishers.

Mason, W. M., Wong, G. Y., y Entwistle, B. (1984). "Contextual Analysis Through the Multilevel Linear Model." En: Leinhardt (ed.) *Sociological Methodology*, 1983-84. San Francisco: Jossey-Bass: 72-103.

May, J. (1990). *Real World Considerations in the Development of an Effective School Incentive Program*. ED 320 271.

Millman, J. (ed.) (1981). *Handbook of Teacher Evaluation*. Beverly Hills, CA: SAGE.

Murnane, R. S. (1991). The Case for Performance-based Licensing. *Kappan*, 73-2: 137-142.

Oakes, J. (1989). What Education Indicators? The Case for Assessing the School Context. *Educational Evaluation and Policy Analysis,* 11: 181-199.

Olson, G. H., y Webster, W. J. (1986). *Measuring School Effectiveness: A Three-year*

Study. San Francisco, CA: American Educational Research Association.

Pollard, J. (1987). *Viewpoints from Selected States on Accreditation and Accountability*. Austin, TX: Southwest Educational Development Laboratory.

Provus, M. (1971). *Disciplinary Evaluation*, Nueva York: McCruthan.

Rosenberg, B. (1973). *Linear Regression With Randomly Dispersed Parameters*. Biometrika, 60: 61-75.

Saka, T. (1989). *Indicators of School Effectiveness: Which Are The Most Valid And What Impacts Upon Them?* San Francisco, CA: American Educational Research Association (ERIC ED306277).

Sanders, W. L., y Horn, S. P. (1995). *The Tennessee Value-Added Assessment System (TVAAS): Mixed Model Methodology in Educational Assessment,* in Shinkinfred, A. J., and Stufflebeam, D. L., Teacher Evaluation: Guide to Effective Practice, Boston: Kluwer.

Scriven, M. S. (1967) The methodology of evaluation. En: R. E. Stake (Ed.). Curriculum evaluation. *AERA Monograph Series on Curriculum Evaluation, Vol. 1*. Chicago: Rand McNally.

Shavelson, R. J., McDonnell, L. M., Oakes, J., y Carey, W. (1987). *Indicator Systems for Monitoring Mathematics and Science Education*. Santa Monica, CA: Rand Corporation.

Stake, R. E. (1967). The countenance of educational evaluation. *Teachers College Record*, 68: 523-540.

Stufflebeam, D. L. (1966). A depth study of the evaluation requirement. *Theory into Practice*, 5, (June): 121-134.

Stufflebeam, D. L. (1996*). "Evaluating School Districts, Programs, and Personnel: Toward a Unified Approach*." Trabajo presentado en la quinta reunión anual del National Evaluation Institute, Bethesda, Maryland, July, 1996.

Stufflebeam, D. L. (1967) The use of and abuse of evaluation in Title III. *Theory into Practice*, 6, (June): 126-133.

Stufflebeam, D. L., y Webster, W. J. (1980). "An Analysis of Alternative Approaches to Evaluation:, *Educational Evaluation and Policy Analysis*, 3, 2: 5-19.

Stufflebeam, D. L., et al. (1971). *Educational Evaluation and Decision-making*, Itaska, IL: Peacock.

Town, S. W. (1973). Action Research and Social Policy. *Sociological Review*. 12, 4: 128-137.

Tyler, R. W. (1949). *Basic Principals of Curriculum and Instruction*. Chicago: University of Chicago Press.

Webster, W. J. (1988). "The Practice of Evaluation in the Public School." En: T. C. Dunham, ed., *The Practice of Evaluation*. University of Minnesota Press: 24-55.

_____. (1997). *Rewarding Effective Schools-Theory and Practice in an Outstanding Schools Awards Program*. Chicago, IL: AERA.

_____, y Olson, G. H. (1988). *A Quantitative Procedure for the Identification of Effective Schools*. Journal of Experimental Education, 56: 213-219.

_____, Mendro, R. L., y Almaguer, T. (1994). Effectiveness Indices: A Value Added Approach to Measuring School Effect:, *Studies in Educational Evaluation*, 20: 113-145.

_____, Mendro, R. L., Bembry, K., y Orsak, T. H. (1995). *Alternative*

Methodologies for Identifying Effective Schools. Distinguished Paper Session, American Educational Research Association, San Francisco, CA, ERIC EA 027 189.

_____, Mendro, R. L., Orsak, T. H., y Weerasinghe, D. (1996). *The Applicability of Selected Regression and Hierarchical Linear Models to the Estimation of School and Teacher Effects*. Nueva York: AERA.

_____, Mendro, R. L., Orsak, T. H. y Weerasinghe (1997). A Comparison of the Results Produced by Selected Regression and Hierarchical Linear Models in the Estimation of School and Teacher Effect. Chicago, IL:AERA. *Multiple Linear Regression Viewpoints* (en imprenta).

Figura 5. Indicadores formativos y sumativos disponibles en las escuelas

Figura 5. Indicadores formativos y sumativos disponibles en las escuelas (cont.)

Figura 5. Indicadores formativos y sumativos disponibles en las escuelas (cont.)

Figura 5. Indicadores formativos y sumativos disponibles en las escuelas (cont.)

Figura 5. Indicadores formativos y sumativos disponibles en las escuelas (cont.)

Figura 5. Indicadores formativos y sumativos disponibles en las escuelas (cont.)

Figura 6. Descripción de variables y metodología utilizada en los índices de eficacia de la escuela

Figura 6. Descripción de variables y metodología utilizada en los índices de eficacia de la escuela (cont.)

Figura 6. Descripción de variables y metodología utilizada en los índices de eficacia de la escuela (cont.)

Figura 6. Descripción de variables y metodología utilizada en los índices de eficacia de la escuela (cont.)

Figura 6. Descripción de variables y metodología utilizada en los índices de eficacia de la escuela (cont.)

Figura 6. Descripción de variables y metodología utilizada en los índices de eficacia de la escuela (cont.)

CAPÍTULO 10

EVALUACIÓN EN LAS REFORMAS EDUCATIVAS DEL ESTADO DE PARANÁ, BRASIL

María Teresa de la Fuente, en colaboración con Heloisa Luck y Corinna Ramos

Una de las reformas educativas más avanzadas en el Brasil es la del Estado de Paraná. Este caso ilustra cómo se han vinculado las actividades de evaluación a la ejecución de las nuevas políticas, buscando desarrollar innovaciones que ayuden a las escuelas a consolidar sus sistemas de administración y, al Estado de Paraná a reorientar sus decisiones e inversiones.

INTRODUCCIÓN

El Brasil está comprometido en una reforma sustancial de su sistema educativo. Este trabajo presenta, en forma sucinta, el contexto de la reforma y describe las iniciativas de cambios que se están ejecutando en los diferentes estados del país. Con el propósito de ilustrar algunas de estas iniciativas, se introducen en forma particular, las que están en proceso de implantación en el Estado de Paraná, con especial atención a las innovaciones que, para su seguimiento y evaluación, han surgido recientemente. Finalmente, se discuten las lecciones que emergen de la corta historia de la evaluación de la reforma educativa en el Paraná.

CONTEXTO DE LA REFORMA EDUCATIVA EN BRASIL

Desde la década de 1980, han venido ocurriendo en el Brasil intensos debates sobre la necesidad de reformar la educación en el país. Estos debates, caracterizados por una tendencia hacia la democratización de la educación, han dinamizado una serie de acciones que, impulsadas por diferentes sectores del país, están definiendo la reforma educativa en todos los niveles del sistema educativo. Las primeras acciones resultaron en casi una universalización de la oferta lográndose una cobertura de la gran mayoría de la población en edad escolar. Esta universalización ocurrió gradualmente, constituyéndose en un proceso de transición largo que ha enfrentado a los sistemas educativos del país con desafíos y problemas tales como:

a) la escasez de profesores cualificados, debiendo contratarse a aquéllos que no siempre tienen la formación adecuada para los cargos que se les asigna,
b) los problemas de gestión derivados de la necesidad de expandir los servicios para cubrir el incremento de la demanda y,
c) el agotamiento de los recursos físicos, financieros y humanos, creando deficiencias que impiden proveer una educación de calidad.

Estos problemas se reflejan en las altas tasas de repitencia de los alumnos, lo que ha generado a su vez una disparidad en la relación edad/curso en el sistema en general; un bajo rendimiento escolar traducido en puntajes deficientes en las pruebas estandarizadas a nivel nacional en los últimas años; y una alta deserción escolar, ya que, "de cada 100 alumnos, apenas 28 completan la enseñanza de primer grado" (Luck, 1996).

En el sistema centralizado de educación del Brasil, todas las políticas educativas son dictadas por el Ministerio de Educación (MEC), que tiene como función la "coordinación de la política de educación, articulando los diferentes niveles y sistemas y ejerciendo una función normativa, redistributiva y suplementaria en relación a las demás instancias educativas" (Articulo 8, párr. 1, Ley 9394/96).

En la actualidad, el MEC articula sus funciones a través de tres organismos: el Consejo Nacional de Educación, el Consejo Nacional de Secretarios de Educación (CONSED) y la Unión Nacional de Dirigentes Municipales de Educación (UNDIME). Estos organismos fueron creados respondiendo a la necesidad de dinamizar la reforma educativa y asegurar un desarrollo integrado del sistema educativo del país.

El Consejo Nacional de Educación se creó recientemente por medio de la Ley 9193 del 24 de noviembre de 1995. Tiene funciones normativas y de supervisión sobre varios niveles del sistema y representa segmentos organizados de la sociedad en sus diversas Cámaras de Educación. El Consejo Nacional de Educación substituye al antiguo Consejo Federal de Educación.

El Consejo Nacional de Secretarios de Educación (CONSED), entidad de derecho privado sin fines lucrativos, fue establecida en septiembre de 1986. El CONSED está constituido por los veintisiete Secretarios de Educación de los Estados y del Distrito Federal. Sus finalidades son: "funcionar como órgano permanente de coordinación y articulación de los intereses comunes de las Secretarías de Educación del país; participar en la elaboración e implantación de políticas nacionales de educación; estimular y promover el desarrollo cuantitativo y cualitativo del sistema de educación pública en el país; y establecer una interacción positiva con todos los segmentos de la sociedad política y civil, con vistas al establecimiento de relaciones sociales justas y de igualdad en un contexto democrático de gestión" (Estatuto Social del CONSED, 1995).

Con una función similar al CONSED, pero a nivel de municipios, se encuentra la

Unión de Dirigentes Municipales de Educación (UNDIME), constituida por todos los dirigentes de educación de los municipios brasileños organizados bajo las cinco regiones geográficas del país: Nordeste, Noroeste, Sudeste, Sur y Centro-oeste. El enfoque principal de la UNDIME es la mejora de la calidad de la educación, y, concretamente, siguiendo el mandato legal, de la enseñanza del primer grado. La estrategia de la UNDIME es la integración entre los municipios para el fortalecimiento común en la búsqueda de resultados.

El esfuerzo articulado de estos organismos ha promovido una mayor comprensión de las formas de intervención necesarias para cambiar la condición actual de la educación brasileña. Las áreas prioritarias de este cambio se encuentran delineadas en el Artículo 214 de la Constitución Federal de 1988 que establece la educación básica obligatoria en el país.

Las directrices de reforma de la educación se aplican sobre el sistema educativo brasileño, el cual está establecido en tres niveles: federal, estatal y municipal. Conforme a la legislación, estos niveles se articulan entre sí en el desempeño de funciones complementarias y convergentes. El MEC dicta políticas y coordina las acciones a nivel nacional para el desarrollo de la educación básica. También ejecuta acciones destinadas a la mejora de la educación superior y profesional. La responsabilidad de proporcionar la educación básica pertenece a los estados y municipios, siendo los estados los encargados de coordinar este proceso. Esta coordinación representa en sí misma un gran esfuerzo de innovación en la educación brasileña ya que, históricamente, se han presentado marcadas diferencias en los procesos educativos, como también en los recursos disponibles en los distintos niveles educativos y en las diferentes regiones geográficas del país.

Tres documentos legales definen en este momento la reforma de la educación en el Brasil, a saber, el Plan Decenal de la Educación para Todos (1993), la nueva Ley de Directrices y Bases de la Educación Nacional (1996) y la Ley 9424 (1996). Estos documentos son el resultado de una amplia movilización de los diferentes segmentos de la sociedad y constituyen las guías más importantes de la reforma educativa nacional.

El primer documento, El Plan Decenal de Educación para Todos (1993-2003) , elaborado con participación de la mayoría de los estados y municipios brasileños, permite "a cada unidad federativa y cada municipio, con base en las directrices y metas del Plan, definir su compromiso y establecer sus propias metas" (Cunha, sd). El Plan establece once líneas estratégicas de acción, "teniendo en cuenta la necesidad de concentrar energías, medios y recursos en la mejora de la educación, propiciando la educación básica completa para eliminar el analfabetismo y la subescolarización de jóvenes y adultos" (MEC, 1993: 45). Estas líneas de acción aprobadas a nivel nacional, son las siguientes:

1. Establecimiento de padrones básicos para la red pública.
2. Fijación de los contenidos curriculares mínimos determinados por la

Constitución.
3. Profesionalización y reconocimiento del magisterio.
4. Desarrollo de nuevos padrones de gestión educativa.
5. Estímulo de las innovaciones.
6. Eliminación de las desigualdades educativas.
7. Mejoría en los niveles de acceso y permanencia en la escuela.
8. Sistematización de la educación continuada de jóvenes y adultos.
9. Producción y diseminación de conocimientos e informaciones en el campo de la educación.
10. Institucionalización de los Planes Estatales y Municipales.
11. Profesionalización de la administración educativa.

El segundo documento, la Ley de Directrices y Bases de la Educación Nacional (LDB), con fecha de 20 de Diciembre de 1996, bajo el número 9394/96, fija las normas de organización y funcionamiento referentes a la educación escolar brasileña, en todos sus niveles y modalidades de enseñanza. Define los niveles del sistema de enseñanza como: educación básica (educación infantil, fundamental, y media); y educación superior. Define también las responsabilidades de los diferentes niveles administrativos: Unión, Estado, Distrito Federal y Municipios, escuela y profesor; caracteriza la educación profesional que debe integrarse a las diferentes formas de educación, inclusive a la escolar; especifica la educación de jóvenes y adultos y la educación especial; y disciplina los usos de los recursos financieros para la educación" (López, 1977: 41).

El tercer documento, nace de la Enmienda 14 de la Constitución que estipula la creación de un Fondo en cada Estado y Distrito Federal. Se trata de *la* Ley 9424, promulgada el 24 de diciembre de 1996, que reglamenta los fondos que deben ser asignados para la educación en los estados y municipios. Esta ley tiene un gran impacto en el alcance de los objetivos de la reforma ya que trata de la redistribución y el traspaso de recursos financieros para el mantenimiento y desarrollo de la educación fundamental, así como también para la valorización del magisterio de este nivel educativo, permitiendo de este modo solucionar los desequilibrios financieros entre localidades.

INICIATIVAS ESTATALES QUE RESPONDEN A LAS DIRECTRICES DE LA REFORMA EDUCATIVA

Actualmente, el sistema educativo brasileño se enfrenta a la urgente necesidad de elevar los niveles de desempeño escolar. El problema del bajo rendimiento escolar afecta a un gran número de países latinoamericanos. De ellos, Brasil se encuentra entre los que exhiben los rendimientos más bajos, junto a Haití, la República Dominicana, Guatemala, el Salvador y Honduras (Wolf, Schiefelbein y Valenzuela, 1994). Como "los sistemas estatales de educación son responsables de cerca de dos tercios de las matrículas de enseñanza fundamental y media en el Brasil,... las Secretarías de Educación tienen un papel importante en la mejora de la educación pública" (CONSED, 1996: 64). Estas mejoras están siendo ejecutadas por medio

de la implantación de proyectos innovadores en los diferentes sistemas estatales. "Se considera que experiencias innovadoras son aquellas que introducen algún tipo de cambio en una determinada cultura y/o práctica escolar, a través de una intervención intencionada o con un propósito previamente delimitado. La innovación debe ser deliberada, planeada y realizada en una secuencia lógica que responda a un propósito previamente delimitado" (Pacheco Mendes, en Leornados y otros, 1994: 10).

Estas innovaciones se agrupan bajo las siguientes áreas de enfoque: mayor coordinación entre estados y municipios; descentralización y mayor autonomía de las escuelas; capacitación y valoración de los profesionales de la educación; mejoras de la calidad de la educación y gestión escolar; introducción de varias estrategias para atender diferentes clientelas; e incorporación de tecnologías de la educación.

El Cuadro 3 (al final del capítulo) presenta los proyectos y acciones innovadores que se están ejecutando en cada Estado del país. El cuadro identifica la "realización de programas que van desde la implantación de cajas escolares como instrumentos de construcción de la autonomía de la escuela, hasta la realización de reformas radicales en la organización de la enseñanza" (CONSED, 1996).

Paralelamente a los esfuerzos de cambios, los estados demuestran un incremento en el interés por introducir prácticas de seguimiento y evaluación para identificar los efectos de la implantación de estas innovaciones. Es importante destacar que esta práctica es reciente en el Brasil, habiéndose iniciado en forma oficial en 1986, cuando fue implantado el SAEB. Sin embargo, ya a partir de 1990, la sistemática de evaluación ha incorporado "factores conceptuales y escolares, que inciden en la calidad de Enseñanza Básica administrados a las diversas poblaciones y subpoblaciones del país (MEC/SEDIAE/DAEB, 1996: 20).

EVALUACIÓN NACIONAL DEL RENDIMIENTO ESCOLAR (SAEB)

El Sistema Nacional de Evaluación de la Educación Básica (SAEB) evalúa el rendimiento escolar de los alumnos y tiene como objetivo obtener información sobre el estado de la educación en el país. Los primeros esfuerzos de evaluación del rendimiento escolar comenzaron en la década de 1980 a raíz de los requerimientos de las organizaciones internacionales que exigían como condición a sus préstamos la organización de una estructura de evaluación. El SAEB aplicó pruebas en 1990 a alumnos de primero, tercero, quinto y séptimo curso de todo el país mostrando que el rendimiento escolar era extremadamente bajo. Las medidas de aprovechamiento de los estudiantes variaban del 30 al 56 por ciento dependiendo de la región y el área curricular evaluada. En algunos casos, "menos de un alumno en 1000 conseguía dominar los contenidos mínimos de los cursos en que estaba" (Nova Escola, 5/97).

El SAEB se aplicó nuevamente en 1993 y 1995. Los resultados obtenidos en estas

dos últimas aplicaciones corroboran el bajo rendimiento escolar del sistema en general y, especialmente, de los cursos más altos (quinto y séptimo curso). Los resultados también muestran que los puntajes que se ubican por encima del promedio corresponden a las regiones geográficas Sur, Sur-Este y Centro-Oeste y que los puntajes ubicados por debajo del promedio están concentrados en las regiones Norte y Nordeste del país.

ESTADO DE PARANÁ: EVALUACIÓN DE LAS ACTIVIDADES DE LA REFORMA EDUCATIVA

El Estado de Paraná, una de las 27 unidades federativas del Brasil, está situado en la región sur del país. Su economía está pasando rápidamente de agrícola a industrial. Cuenta con 8,4 millones de habitantes de diversos grupos étnicos. Según datos de noviembre de 1996, el sistema público de educación estatal está formado por 1.274. 767 alumnos matriculados en la educación básica, distribuidos en 2.046 unidades escolares. En los municipios, el contingente de alumnos es de 545.752 distribuidos en 6.824 escuelas, de las cuales 1.454 se hallaban localizadas en la zona urbana y 5.370 en la rural.

La red pública estatal consta hoy de 44.521 profesores para atender la demanda de educación básica en todos sus niveles y modalidades. De este total, 38.720 actúan como docentes, en el aula; 5.801 son profesionales de la educación que proporcionan apoyo administrativo y pedagógico.

Los datos a seguir revelan que, en relación a los otros estados, el de Paraná se encuentra bien situado con respecto al aprovechamiento escolar de sus alumnos. El aprovechamiento escolar es verificado por el SAEB, con pruebas distintas para los cursos cuarto y octavo de la educación fundamental y tercero de la educación secundaria. El Paraná se situó en 1996, en Lengua Portuguesa (lectura), en el segundo lugar del país con un porcentaje de 52,4 en el cuarto curso de educación fundamental. Esto representa tres puntos porcentuales arriba del promedio nacional. En Matemáticas, en el mismo curso, con un porcentaje del 31,2 por ciento , se sitúa en tercer lugar del país, 1,7 puntos por encima de la norma media nacional. En cuanto al aprovechamiento escolar en el curso octavo de educación básica, Lengua Portuguesa (lectura) obtiene el 67,4 por ciento, ocupando el sexto lugar en comparación con otros Estados, con una diferencia del 1,5 por ciento puntos por encima de la media nacional. En Matemáticas, la posición se mantiene en el sexto lugar, con un aprovechamiento de 36,2 por ciento, lo que corresponde al 0,4 por ciento sobre el promedio nacional.

También con datos del SAEB, la evaluación del tercer curso de educación secundaria en Lengua Portuguesa, en 1996, revela un aprovechamiento de 66,7 por ciento, lo que corresponde al séptimo lugar en el contexto nacional. En Matemáticas, en el mismo curso evaluado, el aprovechamiento escolar es de 36,7 por ciento y la posición relativa a los demás estados es de sexto lugar.

Aun cuando los resultados para el Estado de Paraná no sean los más bajos del país, la Secretaría de Educación de Paraná (SEED-PR) está realizando innovaciones encaminadas a mejorar la calidad de la educación básica. Estas innovaciones se agrupan en tres directrices presentadas en el Plan de Acción de la Secretaría de Educación - Gestión 1995-1998. Se trata de las siguientes: "el alumno permaneciendo con éxito en la escuela y viviendo nuevas y significativas oportunidades educativas; buenos profesores desenvolviendo sus competencias en los ámbitos profesional, personal y cultural, con sistematización y continuidad; y una comunidad participando efectivamente de las decisiones junto al sistema para el alcance de los objetivos educacionales" (SEED-PR, 1995).

Estas directrices exigen la búsqueda de soluciones que (a) eleven los niveles de aprovechamiento escolar de los alumnos; (b) establezcan prácticas de gestión en la escuela y comunidad, y (c) implanten un sistema permanente de perfeccionamiento y actualización de los maestros y profesionales de la educación. Apoyando estas directrices, la SEED-PR ya comenzó un proceso de ejecución de medidas de seguimiento y evaluación que en este momento se encuentra en etapas de desarrollo y en su segundo año de operación.

SEGUIMIENTO Y EVALUACIÓN DEL RENDIMIENTO ESCOLAR EN PARANÁ

En el Estado de Paraná, la evaluación del rendimiento escolar es realizada en forma diferente a la práctica del MEC. Todas las escuelas estatales y, por agregación, las municipales y particulares están cubiertas. Se adopta el principio de la universalidad, en contraste con el de muestra, con el fin de integrar a todas las escuelas en la práctica de seguimiento y de evaluación de la calidad de la educación. Además, se pretende dar la oportunidad a cada profesor, escuela y comunidad del Estado, de situarse en el conjunto de datos globales del sistema estatal promoviendo de esta forma su autoevaluación.

La Evaluación del Sistema de Educación de Paraná forma parte del componente de desarrollo institucional del Proyecto Calidad de Enseñanza Pública de Paraná (PQE), que está cofinanciado por el Banco Mundial y el Gobierno del Estado de Paraná. Comprende la evaluación del rendimiento escolar en la educación fundamental y media. El programa tiene como objetivos los siguientes:

- proveer a la Secretaría del Estado de Educación de informaciones actualizadas sobre el rendimiento escolar,
- conocer el desempeño de la escuelas en cuanto a su organización, gestión, y articulación con la comunidad; y
- promover el desenvolvimiento de una práctica de seguimiento y evaluación de la gestión escolar y pedagógica orientada hacia la mejora de la educación en el Estado.

La meta final de este programa es el establecimiento de un sistema flexible de seguimiento y evaluación en la cual los usuarios puedan utilizar directamente los datos e informaciones obtenidos para su autoevaluación y desarrollo, y para definir acciones educativas adecuadas, en los ámbitos local , regional y central.

La evaluación se lleva a cabo como proyecto colectivo que demanda esfuerzo y en el cual participan alumnos, padres, directores, equipos técnico-pedagógicos de la escuela y representantes de los Núcleos Regionales. Todos ellos actúan bajo la coordinación central de la Secretaría de Educación.

Los padres participan como monitores de la aplicación de las pruebas con objeto de que no sólo contribuyan a la evaluación, sino también y sobretodo, para que pasen a valorar la evaluación como un instrumento fundamental para verificar lo que los alumnos están aprendiendo. Históricamente se ha considerado la evaluación como un instrumento de poder en manos de los profesores por medio del cual se clasifica a los alumnos para un curso posterior, o retenidos en el mismo curso en que se encuentran. Los núcleos regionales, órganos de administración intermedia entre la Secretaría y las escuelas, a través de los miembros de sus equipos técnicos, actúan como coordinadores regionales en el proceso.

Cursos y disciplinas evaluadas

Los cursos y disciplinas se evalúan conforme al programa de ejecución que se viene cumpliendo y supervisando con carácter universal (Cuadro 1).

Cuadro 1. Programa de ejecución de la evaluación. Cursos y disciplinas evaluadas

Paralelamente a la aplicación de las pruebas de rendimiento escolar, se están aplicando cuestionarios sobre la práctica pedagógica de los profesores, de la gestión de la escuela y su organización y articulación con la comunidad. Estos cuestionarios están dirigidos a la dirección , al equipo técnico-pedagógico y a los docentes de las escuelas. Por medio de ellos se planea verificar la relación entre la variación de los índices de aprovechamiento escolar y la naturaleza y sistematización de las prácticas cotidianas en la escuela.

Elaboración de las pruebas

Las pruebas son elaboradas por profesores de la Red Estatal de Educación con experiencia en los cursos objeto de la evaluación. Los profesores cuentan con el apoyo de representantes de los Departamentos de Educación de Primero y Segundo Grado de la SEED-PR, conforme a los cursos que se estén evaluando, los cuales actúan bajo la coordinación de un consultor oriundo de una Institución de Educación Superior (IES).

El proceso de elaboración de las pruebas tiene como objetivos: garantizar la representación del currículo enseñado en el contenido de las pruebas y promover la integración de los profesores dentro de este proceso. Con ello se pretende lograr mayor aceptación de las pruebas por parte de los profesores y una utilización de los resultados por los mismos.

Corrección de las pruebas

La corrección de las pruebas es realizada en cada escuela por un equipo de profesores y aplicadores y por monitores de evaluación. Todos ellos son especialmente designados y capacitados para este fin. Los procedimientos de esta fase se explican con detalle en el Manual de Corrección de Pruebas, y están en la escuela bajo la responsabilidad del director, que recibe una formación específica para tal fin. La opción por la corrección en la propia escuela fue establecida con el objeto de dar información inmediata a la escuela, así como también para desarrollar la participación de los profesores en el proceso de seguimiento y evaluación.

Resultados y usos para mejorar el rendimiento escolar

Los resultados de la evaluación del rendimiento escolar obtenidos en 1995 y 1996 han indicado áreas críticas semejantes a aquellas reveladas por el SAEB. Cada escuela recibe los resultados de cada disciplina evaluada permitiendo identificar los aspectos del aprendizaje en que hay fallas y limitaciones graves. De esta manera, cada escuela puede determinar las intervenciones que se precisan para corregir deficiencias. Se espera que las evaluaciones futuras del mismo curso permitan supervisar y evaluar, por comparación, la efectividad de estas intervenciones. El resultado de estas pruebas permitirá también que el programa de capacitación de la SEED-PR delinee sus programas de acción con base en aquellos aspectos que

requieren fortalecimiento.

SEGUIMIENTO Y EVALUACIÓN DE LA CAPACITACIÓN PERMANENTE DE LOS PROFESIONALES DE LA EDUCACIÓN

Desde 1995, la capacitación permanente de los profesionales de la educación de la SEED-PR se viene desarrollando bajo dos líneas de acción: la de Actualización Continua y la de Perfeccionamiento Teórico-Práctico. Estas líneas de acción están presentes en la programación de la Universidad del Profesor, creada como "institución de apoyo a la educación, con la finalidad de concebir, desarrollar y ejecutar actividades ligadas a la capacitación del personal docente y de apoyo" (Gobierno del Estado/SEED-PR, 1995, Estatuto Social de la Universidad del Profesor, Art. 2) mediante asociaciones con otras instituciones públicas o privadas. Creada el 20 de octubre de 1995 y constituida oficialmente el 28 de octubre del mismo año, inició sus actividades en 1996 apoyada por los recursos financieros del Proyecto Calidad de Enseñanza Pública de Paraná (PQE).

El programa "Universidad del Profesor" integra en sus actividades de capacitación, contenidos y metodología de educación formal y no formal, destinados a promover la actualización personal, como también, la educación en servicio de los profesionales de la educación que actúan en la educación pública y municipal del Estado. La capacitación no es obligatoria. Todos los profesionales del sistema son invitados a participar y lo hacen respondiendo a su motivación y necesidades profesionales. Los principales objetivos de la Universidad del Profesor son:

- propiciar oportunidades a los profesores para el desarrollo de sus competencias en los ámbitos personal, profesional y cultural con el fin de prepararlos para una eficiente actuación individual y colectiva;
- ejecutar estrategias de gestión compartida en la conducción del proceso de actualización continua y calificación teórico-práctica de los profesionales de la educación;
- estimular, dentro de un ambiente profesional, la diseminación de experiencias innovadoras, que emergen de lo cotidiano en el ambiente escolar, con el objetivo de diseminar e implantar aquéllas que puedan ofrecer soluciones a los problemas de la educación en el Estado de Paraná;
- formar a los profesionales de la educación en las áreas específicas del currículo y también en la sistematización de las experiencias pedagógicas y de gestión escolar.

El programa de capacitación de la Universidad del Profesor comprende tres modalidades de capacitación: seminarios, cursos y capacitación en la escuela. Se ofrecen dos tipos de seminarios: Seminario de actualización y motivación— denominado también de "educación avanzada"— y Seminario de actualización curricular, gestión escolar y de apoyo a las prácticas curriculares. En la categoría de cursos, las ofertas se agrupan también en dos tipos: Cursos de especialización y

Cursos de extensión con asesoría técnico-pedagógica. Dentro de la tercera modalidad se agrupan cuatro formas de capacitación: Grupos de estudio, teleaulas, encuentros pedagógicos y módulos de actualización continua.

Entre las modalidades de capacitación en la escuela, cabe destacar a los grupos de estudio (en los que los participantes, procedentes de una o varias escuelas, deciden el tema y la frecuencia de las sesiones de estudio), las teleaulas (realizadas en las diferentes telesalas distribuidas en las escuelas y Núcleos Regionales (NR) del Estado, utilizando técnicas de educación a distancia, videos y, en el futuro, Internet), los encuentros pedagógicos (eventos no regulares, para el tratamiento de temas específicos seleccionados con base en las necesidades y demandas de la escuela) y los módulos de actualización continua (llevados a cabo por los Profesores Multiplicadores, para profesores en la escuela y en áreas curriculares específicas).

Equipo de seguimiento y evaluación

Como parte del Proyecto PQE, se creó a nivel central de la Secretaría de Educación, el Equipo de seguimiento y evaluación. Este equipo tiene como función principal la de acompañar al programa de Capacitación de Profesores, al mismo tiempo que comenzar la implantación de una práctica de seguimiento y evaluación en el sistema escolar estatal. El equipo, en operación desde 1995, realizó un seguimiento y evaluación de la capacitación durante 1996 con el objeto de proveer a la SEED con información cualitativa y cuantitativa sobre la capacitación. Esta información trata de responder a tres preguntas: ¿cúales son las actividades que se implantan adecuadamente?, ¿cómo evalúan los participantes las actividades? y ¿cúal es la relación entre la participación en las actividades de capacitación y el comportamiento de los profesores en el aula?.

La información que se obtuvo a través de estas actividades de seguimiento y evaluación formativos se fue utilizando en la planificación de nuevas actividades, de modo que las opiniones y la evaluación de los participantes ayudó a diseñar de nuevo algunas estrategias. A continuación se incluyen comentarios tomados del Boletín (noviembre de 1996) publicado por el Equipo de seguimiento y evaluación:

> Casi la totalidad (92 por ciento) de los participantes que llenaron los cuestionarios de evaluación demostraron un altísimo grado de satisfacción con los Seminarios de actualización y motivación (Educación avanzada). (...) Durante las visitas a las escuelas se buscó determinar si los efectos de este Seminario se traducen en acciones concretas que resulten en beneficio de los alumnos. Puede decirse que los beneficios ocurren sobre todo en las áreas personal y de relaciones humanas". Con referencia a los Seminarios en Áreas Curriculares se registró que: "una gran mayoría de los participantes considera que los contenidos fueron bien presentados y de acuerdo con la propuesta curricular (86 por ciento, término medio, alcanzando 95 por ciento en el área de Matemáticas).

En este momento todos los niveles del sistema educativo participan en el proceso de seguimiento y evaluación de la capacitación, comprendiendo funciones que van desde la recolección hasta el análisis de datos y la divulgación de resultados. Se pretende implantar un proceso descentralizado, colectivo y realimentador de la práctica que permita realizar un seguimiento cualitativo y cuantitativo de la capacitación.

SEGUIMIENTO Y EVALUACIÓN DE LA GESTIÓN EDUCATIVA: EN BUSCA DE LA EXCELENCIA EN LA ESCUELA

> Una institución educativa orientada hacia el desarrollo integral de los alumnos, los educadores y la comunidad se fundamenta en la articulación de escuela y vida, práctica y teoría, conocimiento y trabajo. Tiene como reto avanzar en el desarrollo de programas y actividades que respondan a los intereses y expectativas de la comunidad, en el mejoramiento permanente de sus procesos administrativos, pedagógicos y didácticos, en la constante reflexión sobre sus prácticas y en la interacción dinámica con su entorno. En este sentido, la institución educativa es gestora de su propio desarrollo porque está en situación de reorientar los procesos escolares a partir de la participación y responsabilidad de todos sus actores. (Salazar Ramírez y Quintero Gutiérrez, 1993: 24).

La calidad de la gestión escolar es un factor determinante de la calidad de la enseñanza porque condiciona el funcionamiento de la escuela, tanto en su alcance de los objetivos educativos, como en su articulación con la comunidad. La gestión escolar en busca de la excelencia consiste en un proceso que articula múltiples factores que, directa o indirectamente, componen la vida escolar en todas sus dimensiones. Dada su complejidad, esa articulación exige un conjunto de habilidades por parte de los directores de escuela, las cuales no siempre poseen. Esto sucede porque los directores de las escuelas paraneenses son elegidos por voto directo de los educadores y comunidad externa (padres y alumnos), a partir de apreciaciones no siempre precisas acerca de sus habilidades de gestión.

En el Estado de Paraná, los directores son profesores y son elegidos para un mandato de dos años, con posibilidad de reelección por un período similar. En las últimas elecciones en 1995 se introdujo el requisito de elaboración de un plan de acción que debe ser presentado por el candidato a la comunidad escolar. Aunque la elección corresponda a un avance del proceso de democratización de la escuela pública, no garantiza necesariamente la efectividad de la práctica de gestión ya que los candidatos a cargo no reciben preparación específica para el desempeño de esta función. Con el propósito de contrarrestar esta situación, la SEED-PR desarrolló determinadas líneas de acción, complementarias entre sí. Una de éstas es la capacitación de administradores a través de talleres de actualización sobre gestión escolar. Otra es una acción denominada Proyecto de Excelencia con la creación del Premio de Excelencia en la Escuela.

El Premio de Excelencia en la Escuela

Con este premio, basado en el modelo Malcolm Baldrige (Education Pilot Criteria) de 1995, se pretende incentivar colectivamente al sistema, partiendo de un conjunto de valores y conceptos fundamentales para obtener excelencia en la escuela. Entre estos requisitos se encuentran el desarrollo del sentido y la importancia de la mejora de la calidad de enseñanza y la aceptación de un proceso objetivo de autoevaluación de la escuela. A partir de estos requisitos, se estableció un cuadro referencial de ocho criterios de desempeño, los cuales funcionan como guías orientadoras de la acción gestora, y almismo tiempo permiten la evaluación interna y externa de la escuela.

Los criterios de desempeño son los siguientes: liderazgo, información y análisis, planificación estratégica y operativa, desarrollo y gestión de recursos humanos, gestión de procesos educativos, operaciones y asociaciones, mantenimiento del medio ambiente escolar, resultados del desempeño de la escuela, enfoque en el alumno y satisfacción de las comunidades interna y externa.

Con el fin de ofrecer apoyo a todas las escuelas estatales en una primera fase de implantación y, ya anticipando la extensión a las redes municipales, se estableció un proceso de búsqueda de la excelencia en las escuelas. Este proceso se inició en 1995 con una invitación para que las escuelas, voluntariamente, discutieran sus prioridades y formularan su "proyecto de excelencia" a partir de un cuestionario enviado a todas ellas. De este modo, las escuelas tuvieron la oportunidad de delimitar su propio camino en la búsqueda de la calidad deseada. Se pretendía con ello transformar cada Plano de Acción del director—presentado durante la candidatura—en un proyecto de excelencia, una vez hubiera sido revisado y asumido por la comunidad escolar.

A partir de la identificación por la propia escuela de sus prioridades, iniciativas para atenderlas e indicadores del éxito en la ejecución de sus planes y proyectos, el director y los profesores completan un formulario de autoevaluación y, voluntariamente, se inscriben para el premio.

El proceso de premio no limita el número de escuelas que pueden participar. Una Comisión de Evaluadores externos a la escuela visita las escuelas inscritas y elabora un documento de análisis que se entrega a cada escuela, informando de esa forma la práctica de gestión administrativa-pedagógica a partir del estado en que éstas se encuentran. La capacitación de directores junto con representantes de las Asociaciones de Padres y Maestros (APM) tiene lugar simultáneamente. Los criterios de desempeño previstos—con su respectiva puntuación (en una escala de 1.000 puntos)—son los siguientes: liderazgo (80 puntos); información y análisis (60 puntos); planificación estratégica y operacional (50 puntos); desarrollo y gestión de recursos humanos (120 puntos); mantenimiento del ambiente escolar (70 puntos); resultados del desempeño de la escuela (250 puntos); enfoque en el alumno y satisfacción de las comunidades interna y externa (250 puntos).

La primera fase del proceso de autoevaluación de las escuelas terminó en noviembre de 1996, cuando se cerró el plazo de envío a la SEED-PR de los formularios de autoevaluación completados por las escuelas. En el proceso participaron 1.659 escuelas (82 por ciento del total). La tabulación de los datos revela una gran dispersión de los resultados, ya que los puntajes van desde los 100 ó 200 puntos (1,7 por ciento de las escuelas) hasta los 900 puntos (una escuela). La mayor concentración de puntos se estableció alrededor de los 500, indicando una autoevaluación en torno de la media de los puntos posibles (26 por ciento de las escuelas). En seguida vienen las categorías de los 400 puntos (24 por ciento de las escuelas) y las de 600 puntos (19 por ciento de las escuelas).

Los datos indican, conforme al cuadro siguiente, que el criterio en el cual las escuelas se juzgan mejores es en el medio ambiente escolar. Es también interesante resaltar que los criterios de desempeño final, que se traducen en la dimensión pedagógica, fueron los que recibieron los menores puntajes.

Cuadro 2. Término medio de los puntajes de la autoevaluación de las escuelas a los criterios de desempeño y puntajes relativos al valor máximo posible

Fuente: SEED-PR (1996)

UTILIDAD, FACTIBILIDAD Y RELEVANCIA DE LOS PROCEDIMIENTOS DE SEGUIMIENTO Y EVALUACIÓN EN PARANÁ

La necesidad de evaluar el rendimiento escolar en todos sus aspectos—como también la de medir el progreso que el sistema en general y las escuelas en particular están alcanzando como resultado de los cambios propuestos—ha hecho imprescindible el establecimiento de actividades de seguimiento y evaluación. En el ámbito del sistema educativo brasileño no ha existido en el pasado una tradición que apoye el desarrollo de sistemas integrados de seguimiento y evaluacion. Sólo en los últimos años, e impulsado por acciones tales como la promulgación de nuevas leyes que afectan a la educación, el establecimiento del SAEB y las nuevas iniciativas que promueven una mayor sistematización en la gestión escolar, se ha empezado a percibir con mayor claridad la utilidad y factibilidad de introducir prácticas de seguimiento y evaluación que faciliten la toma de decisiones basadas en informaciones objetivas y actualizadas.

En este momento, existen varios tipos de propuestas dentro de los estados para supervisar y evaluar acciones innovadoras en sus sistemas educativos. Estas propuestas, en su mayoría, están en etapa de desarrollo y ajuste y constituyen los primeros pasos hacia la creación de sistemas más completos e integrados. Sin embargo, y aun cuando los resultados recogidos sean en su mejoría parciales, ya están proporcionado informaciones puntuales que se están utilizando para redirigir acciones durante la ejecución de los programas.

Dentro de la experiencia paraneense se han alcanzado los siguientes logros:

1. **Mayor participación de la comunidad educativa en los procesos de seguimiento y evaluación.** La participación activa en los procesos de seguimiento y evaluación contribuye a la definición de la relevancia de estas acciones en la práctica diaria. "Desde el punto de vista estratégico, uno de los medios más efectivos para asegurar la independencia institucional ... es precisamente el de articular un mecanismo participativo, en el cual no exista una única autoridad evaluadora. En la medida en que todos los sectores de la comunidad educativa sean alternativamente sujetos y agentes de evaluación y participen en los programas que se pongan en marcha, se asegura una independencia real a través del equilibrio de poderes" (Tiana, 1996). La participación activa dentro de los procesos de evaluación se hace evidente en el Paraná en las siguientes actividades: la confección de pruebas de evaluación de rendimiento hechas por profesores que enseñan las áreas curriculares a ser evaluadas, la participación de padres y maestros en la administración de las pruebas, la corrección de las pruebas dentro de la escuela y la distribución de resultados a las escuelas para que ellas decidan qué medidas correctivas aplicar.

Dentro del área de capacitación, también existe participación activa ya que todas las

personas involucradas evalúan las actividades de formación. Todas las modalidades de capacitación han sido supervisadas y evaluadas por el equipo de seguimiento y evaluación de la SEED-PR por medio de cuestionarios y entrevistas a los capacitados y al capacitador; a través de visitas a las escuelas para ver los resultados de la capacitación en la práctica pedagógica; y por medio de grupos de enfoque que permiten obtener diferentes perspectivas de la capacitación. En el proyecto del premio de excelencia hay participación activa desde el momento en que los proyectos son seleccionados, ejecutados y autoevaluados con participación de los directores, los profesores y la comunidad. Los resultados alcanzados hasta hoy en el área de la participación son prometedores. Considerando que tanto en la capacitación como en el premio de excelencia la participación es voluntaria, los porcentajes de participación alcanzaron, en 1996, el 71 y el 82 por ciento respectivamente.

2. **Mayor difusión de informaciones relevantes.** "La producción y difusión de información relevantes sobre el desempeño del sistema educativo es una manera de apostar a la participación y al compromiso con la educación del conjunto de actores involucrados en ella. Implica, además, apostar a incrementar la capacidad de aprendizaje del propio sistema en todos sus niveles: capacidad para saber qué está sucediendo, para innovar y desarrollar estrategias alternativas y para evaluar sistemáticamente sus resultados." (Toranzos, 1996). En las tres iniciativas expuestas, esta difusión ha probado ser útil, factible y relevante para los que en ella participan. Los resultados de las pruebas de evaluación del rendimiento escolar distribuidos a cada escuela permiten que cada una se ubique en relación al contexto del Estado y proponga sus propias medidas correctivas. El Boletín de seguimiento y evaluación producido por el Equipo provee información con respecto a las actividades de capacitación ya ejecutadas y en funcionamiento. Con respecto al premio de excelencia en la escuela, se espera que, cuando se otorgue, sea posible difundir experiencias positivas y proveer reconocimiento a los esfuerzos locales de todas las escuelas y en particular a las que se destaquen.

3. **Utilización e integración de la informaciones.** Dentro del programa de capacitación ha sido posible usar las informaciones recogidas y, "sobre la marcha", modificar algunas actividades para responder a las sugerencias y necesidades de los capacitados. Por ejemplo, los profesores evaluaron las oficinas de trabajo como "la modalidad de capacitación que más contribuyó a su practica en el aula" y que las "palestras y mesas redondas eran de poca relevancia" (Boletín, noviembre de 1996). Con esta información, la programación de la capacitación fue redirigida para incluir más oficinas de trabajo. También ha sido posible utilizar informaciones similares para identificar problemas de infraestructura que prevén una comunicación rápida entre los niveles central, regional y local y establecer soluciones temporales a las mismas. En el futuro, se planea que estas informaciones estén integradas en un sistema y que éste genere una cultura de seguimiento y evaluación dentro de la SEED-PR. Bajo el liderazgo del Departamento de la educación fundamental de la SEED-PR, ya se ha comenzado a investigar sobre la integración de las actividades de seguimiento y evaluación que se están llevando a cabo en diferentes programas de la

Secretaría.

4. La creación de la Universidad del Profesor que ha traído beneficios mayores de los esperados. **El hecho de poder reunir un gran número de participantes en un mismo local ha favorecido no sólo a las actividades de seguimiento y evaluación,** sino que también ha permitido :

a) Estandardizar la capacitación en gran medida ya que los padrones se mantienen uniformes para todos los capacitados. En el pasado, la capacitación era otorgada por diferentes proveedores en diferentes partes del Estado. Esto permitía una gran variación en cuanto a contenidos y calidad de la capacitación.
b) Enfocar la capacitación a áreas específicas para cada grupo. De esta manera, la capacitación se torna relevante y dinámica en cuanto responde a las necesidades del sistema.
c) Crear oportunidades de intercambio profesional entre el cuerpo docente de la SEED-PR. "Con esto no sólo se logra la capacitación en servicio sino que se fomenta el intercambio de experiencias entre los maestros, lo cual es muy deseable ya que la interacción y la comunicación entre pequeños grupos de maestros propician en ellos un cambio de actitud frente a la introducción de nuevas metodologías"(Colbert, 1990).
d) Facilitar las actividades de seguimiento y evaluación de la capacitación ya que es posible recolectar datos sobre experiencias comunes en un gran número de participantes. Un beneficio adicional de la actividad de recolección de datos es que los participantes viven experiencias prácticas que los sensibilizan a las actividades de seguimiento y evaluación.

5. **La creación del Equipo de seguimiento y evaluación para el componente de capacitación.** Este Equipo ha servido de nexo entre los diferentes niveles del sistema, comunicándose regularmente con los niveles local, regional y central. Ha realizado evaluaciones formativas del programa de capacitación permitiendo la introducción de mejoras y reajustes en sus actividades; y ha proporcionado capacitación regularmente a los representantes regionales responsables por el seguimiento de las actividades de capacitación en las escuelas. Durante 1996, el Equipo realizó estas actividades en forma centralizada desde la ciudad de Curitiba, capital del Estado, siendo los representantes regionales los que se trasladaban a la capital. Durante 1997, con el propósito de responder en forma más amplia a la demanda y reconociendo las diferencias regionales, el Estado se dividió en nueve "polos" los cuales sirven como centro de acción. Con esta estructura, el Equipo se moviliza hacia las regiones para servir las necesidades de asistencia técnica, capacitación y seguimiento en los "polos". Los resultados de estas dos estrategias se compararan al final de 1997.

6. **Para implantar un sistema de seguimiento y evaluación eficientes es necesario establecer las siguientes condiciones**: a) Una comprensión clara de los conceptos de seguimiento y evaluación; y de los propósitos que se persiguen con estos sistemas. Supervisar por supervisar, o evaluar por evaluar no se justifican a

menos que estas actividades sean parte de un plan mayor destinado, en este caso, a mejorar la calidad de la educación. b) Una definición clara de las metas a alcanzar y de los indicadores a ser utilizados; además de esto, deben establecerse los criterios cualitativos y cuantitativos que permitan medir si las metas propuestas fueron alcanzadas. c) Una provisión adecuada de recursos y estímulos que motiven y faciliten el alcance de las metas. Esto presupone el diseño de estrategias de implantación flexibles que permitan atender diferencias, introducir cambios y establecer medidas correctivas si es necesario.

d) Una infraestructura que permita la recolección y el análisis de las informaciones en forma eficiente, y .
e) Un plan de diseminación de la información a todo el sistema.

Establecer estas condiciones requiere de tiempo y recursos que no siempre están disponibles o en las cantidades adecuadas, por ello, uno de los principales desafíos para la ejecución de acciones de seguimiento y evaluación tiene que ver con la disponibilidad de recursos físicos y humanos para llevarlas a cabo.

Los criterios para evaluar la utilidad, calidad, relevancia y factibilidad de las acciones de seguimiento y evaluación en el Paraná están hoy, como en otros Estados, en diversas etapas de desarrollo. Se están utilizando metodologías específicas en forma limitada, mas las tendencias actuales apuntan al desarrollo de un sistema completo e integrado en el futuro.

CONCLUSIONES PARA OTROS CONTEXTOS EN BRASIL Y PAÍSES CON DESAFÍOS SIMILARES

Las iniciativas propuestas como parte de la presente reforma educativa en el Brasil han creado un marco legal, financiero y político que proporciona flexibilidad y apoyo para el establecimiento de medidas innovadoras de mejora del sistema educativo brasileño. Los proyectos descritos en el Cuadro 3 reflejan el consenso y la concientización que hoy existe respecto a la necesidad de cambio y a las áreas en las que estos cambios deben efectuarse. Esta convergencia permite que el sistema total se beneficie de los esfuerzos individuales de cada Estado. La iniciativa gubernamental ha sido y continuará siendo una condición necesaria para el surgimiento de respuestas comunes en la búsqueda de la mejora de educación del país. La creación de un clima propicio para el cambio descansa, en gran medida, en el liderazgo que los gobiernos puedan proveer en todos los niveles del sistema educativo.

La implantación de sistemas de seguimiento y evaluación presenta en sí misma un gran reto. Primero, por la falta de comprensión respecto a la función del seguimiento y la evaluación en el ámbito de las organizaciones educativas en general; y segundo, por la falta de recursos humanos e infraestructura para su implantación. Como todo proceso de cambio, se necesita introducir una etapa de sensibilización antes de comenzar a establecer estos sistemas. La creación de un

Equipo de seguimiento y evaluación dentro de la organización ha permitido crear, en el Paraná, agentes promotores para sensibilizar y capacitar en el uso de tales prácticas.

Uno de los mayores desafíos radica en la obtención de recursos humanos y financieros para apoyar eficientemente un sistema de seguimiento y evaluación. La eficiencia es una condición necesaria ante estas circunstancias. Es por lo tanto aconsejable comenzar con programas limitados que puedan servir como experiencias piloto y puedan expandirse dentro del sistema cuando se disponga de los recursos necesarios y las nuevas estrategias hayan sido propiamente probadas.

La integración de las varias acciones de seguimiento y evaluación es un proceso largo y difícil pues requiere un cambio en los procesos de gestión dentro de la organización y cambios en las actitudes de los personas que producen y utilizan la información. Estos cambio se establecen gradualmente mediante ajustes y negociaciones en lo cotidiano de la escuela y requieren por ello, de tiempo para establecerse y transformarse en sistemas completos al servicio de toda la organización educativa.

REFERENCIAS

Brasil. (1995). *Ley nº 9131/95: altera los dispositivos de la Ley nº 4024/1961. Lex: Coletânea de Legislação Federal.* São Paulo: LEX, v. 59: 2042-2045, out./dez.

Brasil. (1996). *Ley nº 9394/96: establece las directrices y bases de la educación nacional.* Diario Oficial da Uniao (secao I), Brasilia, v.89, n. 248: 27833-27841, 23 de diciembre.

CONSED (Conselho Nacional de Secretários de Educação). (1996). *Relatório de Gestão 1995-1996.* Brasilia: autor.

CONSED (Conselho Nacional de Secretários de Educação). (1995). *Estatuto y Regimento.* São Paulo: autor.

Cunha, C. (s.d.). "Do Plano decenal: compromiso e alcance." En: MEC. *O plano decenal* Brasilia, s.d.

Gobierno del Estado/SEED-PR. (1995). *Plano de Ação da Secretaria de Estado da Educação do Paraná 1995-1998.* Curitiba: SEED-PR.

Leonardos, A.C. y otros. (1994). *Estudo de caso aplicado às inovações educacionais: uma metodologia.* Brasilia: INEP.

Luck, H. (1996). "Community Involvement in the Development of Education: The Paraná Case." Conferencia anual de la Comparative and International Education Society, Williamsburg, VA, U.S.A., Marzo de 1996.

Nova Escola. (1997). "Avaliação Permite Cutucar Os Pontos Fracos Das Escolas." *Revista Nova Escola*, Maio/97. Paraná.

MEC. (1993). *Plano Decenal de Educação para todos.* Brasilia: MEC.

MEC/SEDIAE/DAEB. (1996). *Resultados do SAEB/95: Relatório Final.* Brasilia: DAEB

Salazar Ramírez, A. y Quintero Gutiérrez, R. (1993). *Una mirada cualitativa a la institución escolar*. Bogotá: Ministerio de Educación Nacional.

Tiana, A. (1996). "La evaluación de los sistemas educativos." *Revista IberoAmericana de la Educación*: 37-51. Madrid.

Toranzos, L. (1996). "Evaluación y Calidad." *Revista IberoAmericana de Educación*: 63-78. Madrid

Wolf, L., Schiefelbein, E. & Valenzuela, J. (1994). *Improving the Quality of Primary Education in Latin America and the Caribbean*. Washington, DC: World Bank.

Cuadro 3. Tendencias de las innovaciones en los Estados brasileños

Cuadro 3. Tendencias de las innovaciones en los Estados brasileños (cont.)

Cuadro 3. Tendencias de las innovaciones en los Estados brasileños (cont.)

Cuadro 3. Tendencias de las innovaciones en los Estados brasileños (cont.)

CAPÍTULO 11

CONSTRUCCIÓN DE UN SISTEMA NACIONAL DE EVALUACIÓN EN AGUASCALIENTES, MÉXICO

Margarita Mª. Zorrilla Fierro[1]

El Estado Mexicano de Aguascalientes está desarrollando un sistema de evaluación de la educación desde 1992, como resultado del proceso nacional de descentralización. Este artículo describe brevemente sus orígenes, evolución, características actuales y problemas que enfrenta. Analiza también sus primeros resultados y la relación establecida entre las diversas organizaciones estatales, nacionales y comunitarias, en torno de la evaluación del aprendizaje escolar.

INTRODUCCIÓN

El Estado de Aguascalientes tiene apenas cinco años de experiencia en la gestión de la educación, la cual dependía anteriormente del Gobierno Federal de México. Durante este breve lapso ha tenido que enfrentarse al desafío que significa la transición hacia una educación más abierta y participativa y con capacidad para rendir cuentas a la sociedad, que es a la vez su beneficiaria y su principal recurso. El Estado de Aguascalientes se encontró con problemas culturales, burocráticos y técnicos para implantar un sistema transparente de gestión educativa orientado por el conocimiento de sus propios logros y dificultades. En efecto, la evaluación era considerada como una forma de control político; los resultados de las evaluaciones previas sobre el rendimiento del sistema educativo se mantenían confidenciales. No se publicaban ni utilizaban para conocer o mejorar el sistema educativo.

La construcción de un sistema estatal de evaluación de la educación en Aguascalientes sugiere que es posible obtener resultados en un período de tiempo relativamente corto siempre que existan condiciones institucionales y de liderazgo oportunas. La reflexión que sobre este caso hacen sus participantes indica que:

- se deben estimar de antemano los posibles efectos no deseados de una política de evaluación;

- las actividades de evaluación requieren una responsabilidad institucional;
- es imprescindible contar con el apoyo técnico de las universidades y/o centros de investigación;
- las políticas de difusión de la información deben definirse al comienzo;
- los maestros y, especialmente los supervisores, necesitan capacitación;
- la participación de los usuarios (maestros, alumnos y padres) es un elemento clave para el éxito de las políticas de evaluación y
- el aprendizaje de otras experiencias evita errores y ayuda a clarificar las alternativas.

EL ESTADO DE AGUASCALIENTES

Aguascalientes es un estado pequeño ubicado en el centro de la República Mexicana. Su población no llega aún al millón de habitantes. Su economía se ha transformado en los últimos dieciocho años, al transitar de una sociedad rural a una industrial en la que predominan las empresas de la rama metal-mecánica, textil y, muy recientemente, la informática.

Su sistema de educación básica (preescolar, primaria y secundaria) representa el uno por ciento del sistema educativo nacional con cerca de 230.000 estudiantes.

Hasta 1992, la educación en Aguascalientes dependía en su totalidad del Gobierno Federal. Con el Acuerdo Nacional para la Modernización de la Educación Básica, firmado el 18 de mayo de ese año por el Gobierno Federal, los gobiernos de los estados y el Sindicato Nacional para los Trabajadores de la Educación (SNTE), se inicia una nueva etapa del proceso de federalización, en la cual las entidades federativas se responsabilizan de la totalidad de sus servicios educativos.

La ampliación de la educación básica y el proceso de federalización de la educación coincidieron en Aguascalientes con el inicio de una nueva gestión del Gobierno Estatal, en el cual se planteó como una prioridad el tema educativo. Este interés por la educación se explica por la confluencia de varios factores:

- nueva voluntad política del gobierno;
- capacidad local, es decir, profesionales de la educación expertos en investigación y desarrollo, formados en la Universidad Autónoma de Aguascalientes en los últimos 15 años y con vínculos tanto con la comunidad educativa nacional como internacional;
- conocimiento de la realidad del sistema educativo;
- competencia de la autoridad educativa para la concertación y negociación;
- condiciones geográficas favorables de tamaño y acceso fácil y
- características culturales propicias a la innovación.

En 1992 era imperiosa la necesidad de conocer a fondo las condiciones en que el

Gobierno Federal entregaba el sistema de educación básica al Estado. Con esta finalidad se realizó un primer diagnóstico de indicadores de eficiencia y de equidad (IEA-CETE, 1993a), lo que permitió llegar a tres conclusiones: a) a nivel nacional, el Estado se encontraba entre el noveno y décimoquinto lugar (de un total de 32) en relación a los distintos indicadores considerados (Cuadro 1); b) la eficiencia de la educación básica de Aguascalientes se había mantenido estable en los últimos 20 años y c) en el interior del Estado, la situación educativa no era homogénea y se Observaban diferencias significativas entre los niveles educativos, las localidades, los municipios, las zonas rural y urbana, y las escuelas con distinto tipo de financiación (IEA-CETE, 1993b).

Cuadro 1. Indicadores de eficiencia en Aguascalientes: educación primaria[2]

Sobre esta base, se identificaron varios desafíos: articular los tres niveles de la educación básica; ofrecer una educación básica de calidad, es decir, garantizar la cobertura y el acceso a la escuela, asegurar la permanencia en ella, promover aprendizajes relevantes para la vida de los educandos y cuidar que cada alumno egrese de cada grado oportunamente; diseñar estrategias multidimensionales para la solución de problemas; identificar las singularidades de cada escuela bajo la convicción de que la calidad se gesta en cada plantel escolar e incorporar a los maestros, directores y supervisores en la formulación de las soluciones.

ORIGEN DEL SISTEMA ESTATAL DE EVALUACIÓN DE LA EDUCACIÓN[3]

Dos razones principales motivaron el desarrollo del sistema estatal de evaluación: la necesidad ineludible del sistema educativo de responder por los resultados educativos que produce y la de contar con información sobre la calidad de los aprendizajes para actuar en consecuencia. Tanto la medición de distinto tipo de resultados educativos como los juicios de valor que se construyen a partir de la misma representan el insumo fundamental para orientar las decisiones tanto de macro como de micropolítica encaminadas a mejorar la calidad de la educación.

Por primera vez en varios años, se elaboró en la entidad un marco de política educativa denominado Programa Estatal de Educación 1992-1998, que se difundió ampliamente. Los planteamientos acerca de la necesidad de la evaluación y su vinculación con la planificación estratégica del sector se encuentran desde la primera versión de este programa. En nuestro país, y Aguascalientes no es una excepción, la evaluación es vista como un mecanismo de persecución y ajuste de cuentas. En el magisterio de educación básica en particular, la evaluación es percibida como una forma que utiliza la autoridad para "poner en evidencia a los maestros."

¿Qué fue lo que se hizo para sentar las bases que permitieran construir un sistema estatal de evaluación de la educación?. Hay antecedentes que se remontan a 1983, cuando Don Mario Aguilera Dorantes[4], en convenio con la Universidad Autónoma de Aguascalientes, realizó un conjunto de estudios de diagnóstico del sistema educativo. La realización de estos estudios y la decisión de difundir sus resultados—al menos entre los directivos y supervisores escolares—contribuyó a la salida de este ilustre maestro como responsable de la educación básica y normal en el Estado. El ambiente político de esos años no era en absoluto favorable al impulso académico. La Sección I del Sindicato Nacional de Trabajadores de la Educación (SNTE) estaba en manos de un grupo hegemónico más interesado en defender sus canongías que en atender el desarrollo de la educación. La información derivada de estos estudios se guardó hasta que, más de diez años después, el mismo maestro Aguilera—por entonces presidente del Consejo Nacional Técnico de la Educación—publicara un libro con los resultados del mencionado diagnóstico; sin embargo, su distribución nunca llegó a Aguascalientes.

La relación entre la UAA y el Instituto de Educación de Aguascalientes (IEA)[5] se remonta a 1978, cuando la Universidad creó dos licenciaturas en educación, una orientada a la investigación educativa y otra a la asesoría psicopedagógica. Posteriormente, de 1981 a 1983, se desarrolló un programa de maestría para formar investigadores en educación. Para ello se contó con el apoyo financiero y técnico de la Delegación de la Secretaría de Educación Pública en el Estado, institución que precedió al IEA. A partir de 1983 la relación entre la UAA y el organismo estatal se fracturó totalmente y no fue sino hasta 1992, cuando se inició el proceso de descentralización con el nuevo gobierno del Estado, que se convocó a universitarios especialistas en la educación para elaborar el plan de desarrollo estatal. Tanto el Gobernador del Estado como el Secretario de Educación mostraron interés por establecer vínculos con la UAA desde el principio de su administración. Concretamente, el equipo de trabajo para las áreas de desarrollo del sistema educativo proviene en su mayoría de universitarios. Por su parte, el actual rector de la UAA ha promovido de diferentes maneras los vínculos entre la Universidad y el sector educativo, por ejemplo, a través de la realización de estudios en la Universidad acerca de la calidad de la educación y su evaluación.

Al inicio de la nueva administración gubernamental en 1992 se convino también con la misma UAA realizar una réplica de una parte de los estudios de 1983, en concreto la que se refería a la evaluación de los aprendizajes. Este estudio se llevó a cabo en 1993 y los resultados mostraron que los alumnos no habían mejorado respecto a diez años antes y que en algunos grados y materias incluso habían empeorado.

En 1994 y 1995 se realizaron ejercicios de evaluación del aprendizaje escolar en una muestra de escuelas de enseñanza primaria, tanto públicas como privadas. La publicación de los resultados de las escuelas privadas causó incomodidades e incluso tensiones entre estas escuelas y el IEA. En ese momento la autoridad educativa optó por no publicar los resultados de las escuelas públicas pues eran más deficientes que los de las privadas y se juzgó que no era prudente hacerlo por el impacto negativo que esto podría tener en las escuelas públicas, las cuales representan más del 90 por ciento del servicio educativo de ese nivel y que además se encontraban en una situación de deterioro producto de las políticas de expansión del sistema y del descuido de su desarrollo y calidad.

Aunque es menos conocido el trabajo de evaluación del aprovechamiento escolar que se realizó en el marco del Programa de Carrera Magisterial[6], también es éste un antecedente inmediato, que data de 1993. Se trata de un nuevo sistema de escalafón horizontal creado para medir el desempeño profesional de los docentes. Su importancia radica en ser el resultado de un compromiso entre la Secretaría de Educación Pública (SEP) y el Sindicato Nacional de Trabajadores de la Educación (SNTE). Este compromiso deriva de las negociaciones establecidas en el ya mencionado Acuerdo Nacional para la Modernización de la Educación Básica de 1992.

La evaluación de los resultados educativos, fundamentalmente del aprendizaje escolar, también se introduce en el sistema educativo nacional en el contexto de los Programas Compensatorios que se han venido instrumentando en varias entidades de la República Mexicana desde 1991; actualmente está en vigor en 24 de las 32. Aguascalientes no tiene programas de este tipo por considerar que su rezago no es estadísticamente significativo. A pesar de lo anterior, las políticas estatales de evaluación se han visto enriquecidas por la experiencia nacional derivada de los programas compensatorios.

Con estos antecedentes, la autoridad educativa del Estado se planteó dar un salto cualitativo al establecer en el ciclo escolar 1995-1996 la aplicación de exámenes estatales a todos los alumnos que egresaran de sexto de primaria y de tercero de secundaria. Además, mediante muestreo se midieron los resultados de aprendizaje de los grados tercero, cuarto y quinto de primaria y primero y segundo de secundaria. El propósito era el de diseñar y operar un sistema de evaluación de la calidad de la educación que produjera indicadores diversos sobre los resultados educativos, a nivel estatal, de municipio, de zona educativa y de escuela con el fin de identificar necesidades de apoyo y diseñar estrategias de intervención pedagógica para la mejora de la calidad de la educación.

El concepto de calidad de la educación utilizado comprende la garantía de la cobertura del servicio educativo a todos los niños y jóvenes, su acceso y permanencia en la escuela y la adquisición de aprendizajes relevantes para su vida presente y futura, así como la terminación de cada nivel educativo con oportunidad.

La medida del aprendizaje escolar es quizás el componente más importante del sistema, pero no el único, en tanto un sistema de evaluación se alimenta de otro tipo de informaciones cuantitativas y cualitativas de distintos indicadores, que en conjunto permitirán valorar la mejora de la calidad de la educación.

En marzo de 1997 se promulgó la Ley de Educación del Estado de Aguascalientes, en la cual se establece la facultad y responsabilidad del Instituto de Educación de Aguascalientes de evaluar el sistema educativo; textualmente dice:

> **Artículo 18**. El Instituto de Educación de Aguascalientes diseñará y operará un Programa Estatal de Educación que tome en cuenta todos los tipos, niveles y modalidades educativos. Considerará tanto los aspectos cuantitativos como los cualitativos de las actividades educativas presentes y futuras. El Programa Estatal de Educación se apoyará en los diagnósticos y las evaluaciones del Sistema Estatal de Educación. Éste deberá contar con un sistema de información educativa que contenga bases actualizadas de alumnos, maestros y escuelas, así como de la demanda potencial a partir de las dinámicas demográficas y económicas.
>
> **Artículo 19**. Con el mismo propósito y, en concordancia con las acciones y

> competencias de la Secretaría de Educación Pública establecidas en la Ley General de Educación, el Instituto de Educación de Aguascalientes desarrollará y coordinará un Sistema Estatal de Evaluación educativa que atienda a todos los elementos del Sistema Estatal de Educación y las dimensiones e indicadores del concepto de calidad establecidos en esta Ley. Para que la evaluación que se realice sea rigurosa, en contenidos y metodología, se recabará la opinión y apoyo de maestros y profesionales de la educación.
>
> La evaluación y calificación individual de los alumnos, para efectos de acreditación y certificación de estudios, se realizará por separado, de conformidad con la normatividad federal y estatal aplicable.

Al establecer esta regulación, la evaluación de la educación se legitima jurídicamente y adquiere una importancia que está más allá de las administraciones gubernamentales.

EVALUACIÓN DEL APRENDIZAJE EN 1996

Por su importancia, destaca la experiencia de evaluación de 1996. La aplicación de exámenes denominados estatales se realizó con un censo en los últimos grados de primaria y secundaria (sexto y tercero, respectivamente) y con una muestra de escuelas en los grados tercero, cuarto y quinto de primaria y en primero y segundo de secundaria.

Excepto en sexto grado, en donde además se aplicaron exámenes de ciencias naturales e historia, en los demás grados sólo se aplicaron exámenes de español y matemáticas.

Los exámenes se aplicaron de la siguiente manera: a los 21.400 alumnos de sexto grado de primaria; a una muestra de alumnos de tercero, cuarto y quinto grados de primaria (más de 21.000 alumnos); a los 13.800 alumnos de tercero de secundaria; a una muestra del 25 por ciento de las escuelas, a los alumnos de primero y segundo de secundaria (más de 8.000 alumnos); y finalmente, a los alumnos que ingresaron en primero de secundaria en el ciclo 1996-1997 se les aplicó además una prueba de habilidades académicas, denominada Instrumento de Diagnóstico Académico para Alumnos de Nuevo Ingreso a Secundaria (IDANIS).

La realización de esta tarea de evaluación de envergadura estatal se realizó en cuatro etapas: el diseño de los instrumentos, la aplicación de la pruebas, el análisis estadístico de los resultados y la comunicación y difusión de los resultados.

Los instrumentos fueron diseñados por maestros de educación básica especialistas en las diferentes materias, conjuntamente con investigadores de la educación de la Universidad Autónoma de Aguascalientes[7]. La elaboración de los instrumentos de evaluación se realizó con base en los contenidos curriculares del Plan y Programas

de Estudio de la Educación Primaria de 1993. Las pruebas de español y de matemáticas tanto para primaria como para secundaria estuvieron inspiradas en unos instrumentos españoles. Se basaron estrictamente en el plan y los programas del currículo mexicano para cada grado. Estas pruebas constaban de 50 reactivos cada una[8]. Además, en español, se incluyó una pregunta para que los alumnos desarrollaran un texto escrito que permitiera más adelante indagar sobre cuestiones relativas a la escritura. Las pruebas para cada grado tenían reactivos distintos conforme al programa respectivo aunque su estructura en términos de los ejes temáticos del currículo era la misma. Los exámenes de ciencias naturales y de historia en el caso de sexto grado constaban de 15 reactivos cada uno.

La aplicación de las pruebas se realizó durante la segunda y tercera semanas del mes de junio. Se utilizó un mecanismo de rotación de docentes al interior de cada Zona de Educación Básica (ZEB)[9]; el propósito era que ningún maestro aplicara el examen ni en su grupo ni en su escuela. Además, se incluyó un equipo de observadores externos al proceso.

Los resultados se analizaron calculando el puntaje total de cada prueba, en cada grado. Se trabajó a distintos niveles de desagregación y con distintas variables y se realizaron análisis globales con promedios, dispersiones y comparaciones por materia, género del alumno, medio rural y urbano, escuelas públicas y privadas, turnos matutino y vespertino, municipio y zona de educación básica (ZEB).

El trabajo de retroalimentación, iniciado sin esperar a completar la recopilación y el análisis de todos los datos, fue responsabilidad del equipo del área de evaluación del IEA. Para la difusión y comunicación de los resultados se han seguido varias estrategias, a saber, a cada niño de sexto grado de primaria se le entregó una boleta con el puntaje promedio que obtuvo en el examen; se elaboró un documento de circulación interna con los promedios por escuela y materia; la autoridad educativa circuló boletines de prensa; y se presentaron los resultados en reuniones de supervisores y directores, así como en foros locales y nacionales de investigación y evaluación educativas.

ALGUNOS RESULTADOS DE LA MEDICIÓN DEL INDICADOR DE APROVECHAMIENTO ESCOLAR EN LA ESCUELA PRIMARIA

Los exámenes estatales son pruebas basadas en el currículo y que se aplicaron por vez primera en junio de 1996. Las asignaturas evaluadas, como ya se mencionó, fueron español y matemáticas—para todos los grados—además de historia y ciencias naturales para el sexto grado. Los resultados obtenidos sugieren que la mayoría de los estudiantes no logran el rendimiento escolar esperado y que éste tiende a ser superior en español que en ciencias y matemáticas. También se observa un deterioro en el promedio de estos resultados a medida que los alumnos transitan de un grado a otro, particularmente en matemáticas (Figura 1). Se encontraron

diferencias significativas entre los resultados del medio rural y urbano, siendo mejores los de éste último; y entre escuelas públicas y privadas, con ventaja de las segundas sobre las primeras.

Figura 1. Comparación de promedios globales y por asignaturas entre grados[10]

Fuente: Estado de Aguascalientes, (1996)

Las diferencias entre niños y niñas no fueron significativas, como tampoco lo fueron en el caso de las escuelas públicas que funcionan en turno matutino respecto a las que lo hacen en el turno vespertino, o entre las escuelas que se encuentran en el medio rural del municipio de la capital del Estado y las del resto del mismo.

En el caso de sexto grado se observaron las mismas tendencias. Por el conocimiento experiencial que tenemos acerca del funcionamiento del sector educativo, en particular de las ZEBs y de las escuelas, sabemos que las diferencias indicadas anteriormente se deben en gran medida al tipo de atención que reciben los alumnos, maestros, padres de familia y escuelas. Se observa, por ejemplo, que los resultados del municipio de la capital, en el cual se concentra más de la mitad de la población y es la región más urbana del Estado, son relativamente mejores que en el resto de los municipios. Sin embargo, el municipio de Cosío (ZEB "Q"), ubicado en una de las áreas más pobres del Estado, obtiene los resultados inmediatamente inferiores a los de la capital, siendo incluso mejores que los del municipio de Rincón de Romos (ZEB "V"), el cual es próspero económicamente y cuenta con larga tradición magisterial. Esta situación se debe en gran parte a que, en Cosío, el equipo de supervisión está realizando una tarea muy importante, mientras que en

Rincón de Romos no ha sido posible contar con un liderazgo académico sólido. De igual manera se destacan los resultados del municipio de Calvillo (ZEB "O-P"), el cual en 1991 mostraba los peores indicadores de eficiencia educativa; también este despegue de sus resultados de aprendizaje se debe al trabajo del equipo de supervisión.

Estos resultados ponen de manifiesto la necesidad de fortalecer la gestión pedagógica. La identidad de los planteles escolares y de las zonas de supervisión adquieren especial relevancia en las nuevas concepciones de la gestión descentralizada.

Sin embargo, la historia de la supervisión y dirección escolares—concebidas como puesto o cargo, desde su definición normativa y su constitución en la práctica, su inscripción en las lógicas de la burocracia y su particular configuración política—plantea fuertes interrogantes y desafíos para reorientar estas funciones directivas y de supervisión hacia una gestión regida por criterios profesionales.

Los resultados en la educación secundaria no son diferentes en cuanto a sus distribuciones respecto a los de la educación primaria. Los promedios siguen las mismas tendencias de deterioro.

¿Qué otras hipótesis se pueden plantear para explicar estos resultados?. Queremos apuntar algunos argumentos de tipo cultural. El Día del Maestro[11], el Secretario de Educación de Aguascalientes se dirigió a los maestros diciendo:

> ¿Por qué unos estados del país tienen mejores resultados educativos que otros?, ¿será acaso que en este mosaico de culturas que es nuestro país, unas sociedades han confiado en su educación y otras no?. Parece que hay algo. Aguascalientes, a pesar de sus buenas condiciones de tamaño y comunicación, llegó a los noventa con indicadores de desarrollo educativo ubicados alrededor de la media nacional, incluso menos. Me atrevo a decir que la sociedad de Aguascalientes, en el pasado cercano, le atribuyó poca importancia a su educación, no la cuidó. Nuestra industrialización es muy reciente y en consecuencia las exigencias que ésta conlleva también lo son. Mientras nuestra vida era culturalmente rural, no necesitábamos de la educación escolar, las exigencias eran menores, la vida era sencilla. Ahora, los escenarios han cambiado radicalmente y en muy poco tiempo, de ahí la creciente exigencia de que los niños y jóvenes reciban una educación de calidad, que realmente aprendan cuestiones relevantes para su vida presente y futura, como personas y como ciudadanos. Creo que todos estaremos de acuerdo en que la educación escolar es el agente fundamental de la cultura y, si la cultura es el factor que hace la diferencia entre una sociedad y otra, la conclusión parece evidente (Álvarez, 1997).

La hipótesis cultural va más allá, trasciende y permea toda la maquinaria del sector educativo, la formación inicial de los docentes, la actualización permanente de los profesores en servicio, las condiciones de las escuelas, las relaciones de la escuela

con la comunidad y con los padres de familia, los métodos pedagógicos utilizados por los docentes, el valor del conocimiento, el tipo de relación entre los distintos agentes y actores[12] de la educación, las tradiciones culturales de las distintas regiones del Estado y del país, los mecanismos de acceso y permanencia en la supervisión y dirección escolares, la compleja vida sindical y otros factores más. Estas reflexiones sólo suponen el punto de partida para una indagación más a fondo de los *modus operandi* de cada una de estas dimensiones del desarrollo ordinario de la educación escolar. La pregunta está presente: ¿cuáles son las características de la oferta educativa en un determinado lugar y cómo interactúan éstas con las de la demanda y las del contexto para producir un tipo de resultados educativos?.

EFECTOS DE LAS POLÍTICAS SOBRE EVALUACIÓN

Las medidas de políticas ejecutadas en Aguascalientes en relación con la evaluación de los resultados de aprendizaje han sido relativamente exitosas, en el sentido de que no han sido impugnadas ni por la sociedad ni por la organización sindical de los maestros, como ha sucedido tradicionalmente.

Buena parte de este éxito inicial se debe a la voluntad de la autoridad educativa, en este caso la del Secretario de Educación. Lo mismo podría decirse de otros estados del país. Cuando existe voluntad política genuina para atender a la evaluación, se puede perfilar una oferta de información y juicio sobre la educación a pesar de las enormes dificultades que hay que afrontar. Otro factor determinante ha sido la conformación de un equipo técnico de profesionales con experiencia en investigación educativa.

Pero el trabajo en el lado de la demanda es también fundamental. En el caso de Aguascalientes, la difusión de los resultados de los niños de sexto grado de primaria ha sido el otro componente fundamental para la implantación del sistema de evaluación. La autoridad educativa estatal tuvo el valor de publicar y diseminar los resultados de la evaluación. Como se mencionó anteriormente, cada alumno recibió de manera personalizada, mediante una boleta, el resultado de su examen. En la prensa se difundieron los resultados mediante boletines dirigidos por el Secretario de Educación. Los líderes sindicales hablaron en favor de la evaluación y de la necesidad de mejorar la calidad del aprendizaje de los alumnos. El Gobernador del Estado entregó premios a los 100 mejores niños del Estado y, de éstos, diez fueron enviados a saludar al Presidente de la República.

En el verano de 1996, en la reunión de autoridades estatales de educación básica[13], se presentaron los resultados de las pruebas con mayor detalle, destacando las distintas comparaciones. La reacción en primera instancia fue la de negar la realidad que mostraban los resultados, o al menos minimizarla. En lo que respecta al equipo académico que realizó los trabajos de recolección y análisis de los datos, la actitud fue de preocupación al considerar que para mejorar los aprendizajes no basta con garantizar la oferta del servicio, es decir, que la existencia de suficientes escuelas y maestros para atender a la población es necesaria pero no suficiente. Cumplir con el

precepto constitucional de garantizar la educación a todos los individuos es sobre todo garantizar el derecho a aprender, a tener oportunidades de desarrollo. Por esta razón, los actores del sistema educativo con funciones de dirección y de toma de decisiones presentan gran resistencia al proceso evaluativo.

Por otro lado, la aplicación de exámenes para valorar el aprovechamiento escolar de los alumnos es una actividad externa a la escuela pero no al sistema educativo. La instancia responsable de realizar la evaluación es la Dirección de Desarrollo Educativo del propio Instituto de Educación de Aguascalientes. Sin embargo, en 1983, la evaluación no sólo fue externa a la escuelas sino también al sistema educativo mismo; fueron los universitarios quienes la realizaron y nada peor le puede suceder al magisterio que sentir que es puesto en evidencia por otras personas consideradas totalmente ajenas a la problemática del sistema educativo.

A pesar de las limitaciones que impone el no contar con estudios específicos sobre el impacto de estas acciones de evaluación, podemos decir que la sociedad perdió el miedo a informar acerca del estado del aprendizaje de los niños y jóvenes. Los padres de familia, de una u otra manera, conocen el nivel de aprendizaje de sus hijos o al menos lo intuyen. La difusión de los resultados corrobora las sospechas acerca del deterioro de la calidad de la educación impartida en las escuelas.

Colocar en el centro del debate, del análisis y del trabajo mismo el aprendizaje de los educandos es una especie de revolución copernicana porque esto permite ahora que tanto los docentes como los padres de familia y los mismos alumnos persigan otra manera de aprender. La centralidad del educando en el proceso educativo transita paulatinamente del mero discurso a ser una realidad.

En el caso de las reacciones de los docentes, un elemento que a nuestro juicio mitigó el impacto de los bajos resultados fue la inaguración, en enero de 1996, de un amplio programa para la actualización de todos los profesores de educación básica en servicio. Por vez primera en más de 30 años se atendía a los docentes mediante novedosas estrategias de cursos y talleres. De entrada, el impacto político en los profesores ha sido favorable; en sus propias palabras, se sienten atendidos, "tomados en cuenta".

El efecto de la evaluación desde la perspectiva de los docentes es variable. Al continuar con una atención diversificada para cuidar las cuestiones de la formación permanente del profesorado, los resultados de la evaluación son objeto de conversación; sin embargo, el autoanálisis crítico no tiene lugar aún en las discusiones de los docentes. Es recurrente decir que los resultados se deben a cualquier factor del contexto cultural y socioeconómico de los educandos, pero menos a los mecanismos mediante los cuales operan las escuelas, en las que finalmente cada docente se encuentra integrado.
Los resultados de la evaluación del aprovechamiento escolar que se está realizando en este fin de ciclo 1996-1997 podrán aportar información respecto al avance o retroceso de este indicador. Es importante mencionar que, en la administración

central del sistema educativo, tenemos reservas acerca de la mejora de los aprendizajes de los alumnos. Parece que la incorporación de innovaciones como el programa de actualización y el nuevo currículo están exigiendo del docente nuevos conocimientos y habilidades pedagógicas. Es más, éstos requieren de tiempo—no sabemos cúanto—para ser asimilados e incorporados en las prácticas docentes cotidianas. En otras palabras, el discurso y convencimiento intelectual requiere ser procesado de tal manera que permita su apropiación y genere entonces nuevas prácticas pedagógicas a nivel de escuela y de aula.

LOS RIESGOS Y LAS ALTERNATIVAS

Es importante advertir sobre los posibles efectos no deseados o "perversos" de un sistema de evaluación educativa. En nuestra experiencia, en el sistema educativo mexicano, las innovaciones tienden a pervertirse de manera bastante rápida. Algunos de estos efectos no buscados serían los siguientes: que en las escuelas sólo se enseñe para aprobar los exámenes estatales; que la evaluación sea una función que solamente se le atribuya a las autoridades centrales del sistema y entonces los docentes se desentiendan de ella; y que los resultados se utilicen para estigmatizar a los alumnos, para presionarlos irracionalmente, incluso para expulsarlos de las escuelas.

Respecto a los docentes, puede suceder lo mismo: estigmatización del maestro mediante mecanismos de control sin apoyo; que se desate una competencia salvaje entre las escuelas y los docentes; y que perjudique más que favorezca el desarrollo equilibrado e integral de las personas. Por el momento, no sabemos todavía si el currículo es pertinente o no y, sin embargo, la evaluación del aprendizaje de los alumnos se encuentra ya vinculada a la promoción salarial en el marco del escalafón horizontal que promueve el programa de carrera magisterial.

Para poder revertir este tipo de efectos no deseados, hay que trabajar en varias direcciones, a saber:

- Continuar con la medición del indicador de aprovechamiento escolar en los grados de fin de nivel, es decir, en sexto de primaria y en tercero de secundaria y agregar tercer grado de educación media.
- Resolver la concepción y el sentido de un sistema estatal de evaluación de la educación y su ubicación relacional en una nueva manera de plantear la estructura organizativa que transite de estancos aislados a una distribución matricial de las tareas, para que éstas se resuelvan por consenso de todas las instancias responsables de las tareas educativas; en este sentido es necesario colocar los problemas en manos de los interesados para pensar las soluciones y ponerlas en práctica.
- Promover el trabajo de investigación educativa mediante convenios con las instituciones de educación superior que la realizan. La finalidad es la de atender a cuestiones como el diseño y la prueba de instrumentos y de metodologías para la evaluación de centros escolares, la formación y fortalecimiento de grupos de

trabajo de expertos en la administración del sector educativo y la elaboración de estudios evaluativos que contextualicen los resultados según las características de la oferta y la demanda educativas.

- Definir las políticas institucionales para la difusión y diseminación de resultados y, a partir de ellos, orientar y diversificar programas de desarrollo como el de la actualización permanente del profesorado.
- Atender prioritariamente a la capacitación de supervisores, directivos y docentes sobre el tema y tarea de la evaluación educativa, identificando las responsabilidades de cada uno en su ámbito de competencia. La evaluación tiene que convertirse en un instrumento de apoyo, en un mecanismo de rendición de cuentas y en un camino para el crecimiento y el aprendizaje de las personas, tanto de los niños y jóvenes como de los adultos educadores.
- Conocer y aprender de otras experiencias derivadas de la propia investigación evaluativa y de los sistemas de otros países del continente y del mundo. Éstas tienen que convertirse en un referente obligado para no recorrer el camino de los mismos errores.
- Discernir sobre el componente político que necesariamente toda actividad evaluativa conlleva. No es sólo asunto de la voluntad de quien dirige el sistema educativo, sino que se ha de colocar la evaluación en el centro del consenso a alcanzar entre los agentes y actores de la educación. Como señala la última Carta Informativa del IIPE-UNESCO (Ross, 1997), la investigación sobre los resultados educativos y la política en cuanto a la toma de decisiones acerca de dónde y cúanto invertir en educación es una *mixtura compleja*. La enseñanza más importante es la de la necesaria dosis de prudencia que hay que tener al establecer conclusiones de política a partir de los informes de resultados educativos, si éstos no han sido precedidos de estudios detallados sobre el contexto y las condiciones en que éstos se producen.

ENSEÑANZAS DE LA EXPERIENCIA INICIAL

La experiencia brevemente descrita nos enseña que es posible construir sistemas estatales de evaluación. En particular destacamos los factores que la favorecen y que consideramos fundamentales.

En primer lugar, la decisión y el apoyo político del Secretario de Educación del Estado, los cuales se traducen en la convicción de que la evaluación de los aprendizajes es una necesidad; en la definición de que la evaluación debe ser externa a la escuela pero no al sistema educativo; en la asignación de recursos presupuestarios; en la supervisión y seguimiento del proceso y en el acompañamiento del equipo responsable.

En segundo término, la formación de un equipo técnico de profesionales con experiencia en investigación educativa, en la realización de estudios evaluativos, en la docencia en educación básica y en la informática. El equipo estatal está integrado por diez personas de las cuales tres son las que dirigimos los procesos de

elaboración de pruebas y su validación, el diseño de la operación y capacitación para el levantamiento de datos y la elaboración de distinto tipo de informes de resultados para diferentes usuarios. Este grupo de personas ha desarrollado la capacidad de gestión del sistema de evaluación en el sentido de traducir en su diseño y operación los planteamientos nacionales y estatales[14] de política educativa.

En tercer lugar, el valor que tuvo la autoridad educativa estatal para publicar y deseminar los resultados a la sociedad en general y a distintos usuarios dentro del sistema estatal de educación en particular.

Otro factor que, a nuestro juicio, ha sido muy importante es el que se refiere a la actualización de los docentes de primaria y secundaria. Como se mencionó, todos los profesores están siendo atendidos mediante un innovador esquema de formación de profesores en servicio. Esto se realiza a través de cursos-taller—que se llevan a cabo en días hábiles y en horario laboral, con materias y con instrucciones itinerantes— en los que se trabaja con los grupos de profesores que corresponden a una zona de educación básica. Este programa se inició seis meses antes de la primera aplicación de los exámenes estatales.

El quinto factor que identificamos, y que es menos tangible, se refiere a la confianza y respeto que el equipo de personas de la Dirección de Desarrollo Educativo y del área de evaluación ha logrado obtener en los últimos cuatro años a través de las tareas que realiza en la actualización, la investigación, la innovación y la propia evaluación.

Este elemento se puso de manifiesto este año. Por un lado se aplicaron pruebas a los alumnos de tercero de primaria a tercero de secundaria de los maestros que están inscritos en el programa de carrera magisterial. Esta evaluación de aprovechamiento se inserta en un sistema nacional controlado por la misma SEP a nivel central. Desafortunadamente, esta evaluación está muy desprestigiada y ha generado efectos no deseados que han pervertido este mecanismo.

Por otro lado, la mayoría de los docentes percibe lo que llamamos exámenes estatales como algo propio y original de Aguascalientes, que se realiza más para saber cómo estamos en cuanto al sistema educativo que para catalogar o poner en evidencia a los profesores.

Hasta el año pasado (1995-1996), la aplicación de los exámenes del indicador de aprovechamiento escolar de carrera magisterial la realizó otra instancia del Instituto de Educación de Aguascalientes. Este año—en que por cuarta vez se aplican estos exámenes—el Área de Evaluación se hace responsable por vez primera de hacerlo, situación que ha hecho posible identificar problemas, carencias, desviaciones y simulaciones.
Finalmente, en el discurso político del Secretario de Educación del Estado mencionado anteriormente, éste insistió en que la evaluación es una pieza clave para mejorar la calidad de la educación y afirmó que los bajos resultados—atribuibles a

la estructura y al funcionamiento del sistema estatal de educación en su conjunto— consituyen el piso del cual partir. Si bien el aprovechamiento escolar de los alumnos es un reflejo del trabajo del docente, está claro que también lo es del desempeño de los distintos niveles de autoridad: del director de escuela, del supervisor escolar y de las autoridades y mandos medios del Instituto de Educación de Aguascalientes.

NOTAS

[1] Como todos los trabajos intelectuales, éste no puede ser concebido como una elaboración exclusiva de quien lo firma. Lo que aquí se presenta es la integración de varios textos que documentan las políticas de evaluación de la educación en el Estado de Aguascalientes. Destaca el trabajo sobre análisis de resultados de distintas mediciones del aprendizaje escolar realizado por el Mtro.Daniel Eudave y sus colaboradoras. Los comentarios siempre analíticos y críticos de Pilar González, actualmente Directora de Planeación del Instituto de Educación de Aguascalientes han enriquecido este modesto documento. La responsabilidad de lo que aquí se plantea es solamente de la autora.

[2] Como se puede apreciar, estos indicadores de eficiencia no incluyen medición de resultados de aprendizaje.

[3] Zorrilla, F. M. Coord. (1997) *Descentralización e Innovación Educativa. Una mirada desde el proceso de gestión de la Innovación. El Caso de Aguascalientes.* Este trabajo fue financiado por la Fundación Ford y coordinado por El Colegio de México.

[4] Por aquél entonces, era el Delegado de la Secretaría de Educación Pública en la entidad. El maestro Aguilera es todo un personaje en el sistema educativo nacional. Ingresó como maestro rural en 1924 y en dos ocasiones fue Oficial Mayor de la misma Secretaría.

[5] En el Estado, la conducción del sistema educativo es responsabilidad de un organismo público descentralizado, el Instituto de Educación de Aguascalientes. La máxima autoridad es el Director General cuyas funciones son las de un secretario de educación.

[6] Carrera magisterial es un sistema de evaluación del desempeño docente relacionado con un escalafón horizontal, en el cual se valora, entre otros factores, el aprovechamiento escolar de los alumnos. Esta evaluación tiene impacto en el salario de los docentes.

[7] El Departamento de Educación de la UAA ha desarrollado, entre otras líneas de investigación educativa, la evaluación del aprovechamiento escolar tanto en la dimensión de conocimientos como en la de habilidades y cuenta con tres investigadores con experiencia en el campo. Respecto a la evaluación de los aprendizajes realizada en 1996, la UAA participó con personal que trabajó en la elaboración de las pruebas de español y de matemáticas para los grados de tercero, cuarto, quinto y sexto de primaria. Las pruebas de ciencias naturales e historia para sexto de primaria fueron elaboradas por maestros de primaria con asesoría de los investigadores de la UAA. Las pruebas de español y de matemáticas para los tres grados de secundaria fueron elaboradas por docentes de dicho nivel. La reproducción y distribución de las pruebas fueron responsabilidad del IEA.

[8] Los de español y matemáticas resultaron válidos en el sentido de que la mayoría de los ítemes tenían capacidad de discriminación y grado de dificultad aceptable.

[9] Los servicios de educación básica de la entidad se encuentran distribuidos en unidades geográficas denominadas Zonas de Educación Básica (ZEB). Estas zonas están integradas por escuelas de preescolar, primaria, secundaria, servicios de educación física y de educación especial. Son atendidas por un equipo de supervisores, dirigido por un coordinador. En total hay 22 zonas, 12 en el municipio de la capital y una en cada municipio hasta sumar 10; son de distinto tamaño en cuanto a la cantidad de escuelas, docentes y alumnos que atienden. Para mayor información ver el trabajo de Márquez, M. y Zorrilla, M. (1997). *Redefinir la supervisión para atender la escuela singular*. En imprenta.

[10] En este documento se presentan los resultados obtenidos por los alumnos de tercero, cuarto, quinto y sexto de primaria del Estado de Aguascalientes en los exámenes estatales aplicados en junio de 1996. Se escogió esta época del año confiando en que, para entonces, se habrían cubierto la mayoría de los contenidos programáticos de las materias evaluadas. Se hizo un muestreo estratificado para la selección de las escuelas a evaluar en los grados de tercero, cuarto y quinto. El Estado se dividió en estratos según los siguientes criterios: medio urbano y rural.; condición socioeconómica y de bienestar social: favorable, desfavorable y marginal. Para el caso de sexto grado de primaria, se hizo una aplicación censal.

[11] En México, el 15 de mayo, en que se celebra el día del maestro, las escuelas y las ciudades se llenan de fiesta en homenaje a los maestros.

[12] Una distinción entre agente y actor de la educación que puede resultar comprehensiva y fecunda es la siguiente: el actor recibe un libreto y lo actúa, mientras que el agente actúa, decide y construye.

[13] Nos reunimos el Secretario de Educación del Estado, todos los mandos medios del IEA, los coordinadores de las zonas de educación básica y algunos responsables de proyectos.

[14] La Ley General de Educación, el Programa de Desarrollo Educativo 1995-2000 del Gobierno Federal, la Ley de Educación del Estado de Aguascalientes y el Programa Estatal de Educación 1992-1998.

REFERENCIAS

Álvarez G., Jesús. (1997). *Mensaje del Día del Maestro. Mayo de 1997*. Mecanograma.

Estado de Aguascalientes. (1993). *Educación básica. Diagnóstico de indicadores de eficiencia y equidad*. Aguascalientes: Autor.

Estado de Aguascalientes. (1996). *Sistema estatal de evaluación de la educación. Informe de resultados*. Aguascalientes: Autor.

IEA-CETE (1993a). *La educación básica en Aguascalientes. Diagnóstico y propuesta*. Aguascalientes: UID.

IEA-CETE.(1993b). *Programa de Modernización de la Función Supervisora*.

Documento Rector. Aguascalientes: 87.

Ross, Kenneth N. (1997). "Investigación y política: una mixtura compleja." *Carta Informativa del IIPE,* XV, 1: 1-4.

Zorrilla, F. M. Coord. (1997). *Descentralización e Innovación educativa. Una mirada desde el proceso de gestión de la Innovación. El Caso de Aguascalientes.*

DILEMAS Y OPCIONES DE POLÍTICA PARA LA EVALUACIÓN EDUCATIVA

Benjamín Álvarez H. y Ray Chesterfield

La evaluación educativa es un tema de importancia estratégica para las naciones. Tanto las políticas sociales nacionales como los acuerdos internacionales reflejan el interés de todos los países por lograr para sus niños y jóvenes resultados de aprendizaje de alta calidad. El acuerdo realizado en la Conferencia Mundial sobre Educación para Todos celebrada en 1990 en Jomtien, Tailandia, ilustra la creciente demanda por el mejoramiento de los sistemas educativos en el mundo entero. Aunque existe un consenso amplio sobre esta necesidad y se están ejecutando numerosos esfuerzos de reforma educativa, disponemos de poca información que nos permita juzgar su éxito, o adecuar los sistemas educativos a las necesidades de diversos tipos de estudiantes.

Los capítulos anteriores de este libro describen la evolución y las tendencias recientes de la evaluación educativa y discuten las innovaciones introducidas como resultado de las reformas educativas de esta década en América Latina. Estas tendencias sugieren nuevos dilemas y opciones de política que los países, los estados y las familias tienen que enfrentar. Este capítulo destaca algunas de esas tendencias, resume el estado de la práctica y busca contribuir a la discusión de la política.

TENDENCIAS PRINCIPALES

Medición del rendimiento académico

Siguiendo las tendencias internacionales, en América Latina se están multiplicado notablemente los esfuerzos para determinar el nivel de desempeño de sus estudiantes en destrezas básicas tales como lenguaje y matemáticas. El instrumento más generalizado para hacerlo han sido las pruebas de conocimiento. Hasta hace poco, el énfasis se puso en pruebas basadas en normas estadísticas que son muy comunes en los Estados Unidos. Esto se debe en parte a que muchos países no han establecido criterios para el desarrollo de destrezas básicas en los diversos grados o cursos que forman el sistema. En algunos países, la evaluación del aprendizaje para elevar la calidad de la educación se está construyendo sobre la experiencia anterior

de sistemas de prueba diseñados para la selección de candidatos para niveles superiores de escolaridad.

Todos los trabajos teóricos y prácticos presentados reconocen la importancia de utilizar instrumentos adicionales para valorar los diversos resultados del aprendizaje escolar y medir otras variables importantes relacionadas con el rendimiento académico. Algunos incluso describen enfoques alternativos e instrumentos para relacionar el desempeño de los estudiantes con el desarrollo del currículo o de los factores ambientales. Es decir, existe una tendencia que busca no solamente identificar los resultados del aprendizaje logrado en la escuela, sino también explicarlos con el fin de ejecutar intervenciones estratégicas. No obstante, aspectos más amplios del comportamiento de los estudiantes, tales como la formación de actitudes y valores que constituyen objetivos de varios de los programas de reforma educativa, no se incluyen generalmente en los sistemas de evaluación.

Al interés en la medición del rendimiento académico se ha sumado un movimiento por establecer estándares nacionales de desempeño escolar y realizar comparaciones y acuerdos internacionales encaminados a mejorar el rendimiento académico de los estudiantes de cada país, utilizando el aprendizaje obtenido por medio de una interacción sistemática entre las naciones. No obstante, un país necesita de una sólida infraestructura nacional de información e investigación para sacar ventaja de su participación en tales actividades.

Definición de la calidad educativa

Aunque los capítulos anteriores hacen referencia a diversas concepciones de la calidad de la educación, la mayoría de los sistemas descritos parecen definirla en términos de la eficiencia interna del sistema o bajo la perspectiva de la función de producción de la educación, asimilándola a las empresas productivas.[1] El supuesto es que los recursos y procesos escolares establecidos en una escuela o sistema determinan el grado de aprendizaje que se logra. En forma paralela, la combinación apropiada de insumos puede reducir el desperdicio del sistema e incrementar su eficiencia.

Una definición relacionada a la anterior y que aparece generalmente en las discusiones sobre la justificación de la evaluación, es la basada en el concepto de eficiencia externa. Esto es, la capacidad de los productos de un sistema educativo para funcionar y competir en una sociedad y aumentar la habilidad de una nación para participar en la economía global (Heyneman, 1997). Pero, al seleccionar un esquema de insumo-proceso-resultado para orientar las acciones tendientes a mejorar la calidad de la educación, conviene tener en cuenta que se trata siempre de representaciones parciales de procesos en los que influyen numerosos factores.

Los sistemas de evaluación discutidos dedican menos atención a la calidad educativa concebida como la creación de un ambiente adecuado para el desarrollo cognoscitivo y la adquisición de destrezas. Este concepto está basado en un cuerpo

de conocimiento reciente que sugiere que el aprendizaje se logra en un ambiente de colaboración, cuando se construye sobre la base de conceptos familiares y a través de una red de interacciones sociales de apoyo y colaboración que permite a los estudiantes interpretar información, más bien que acumularla (Gardner, 1991 y Levinger, 1996). Dado el énfasis de muchos de los esfuerzos de reforma en enfoques constructivistas del aprendizaje, esta perspectiva para considerar la calidad merece consideración.

Sin embargo, esta tendencia hacia la búsqueda de ambientes educativos más estimulantes y que faciliten logros de aprendizaje de mayor nivel, cubierta bajo la categoría de calidad de la educación está produciendo, más que un concepto abstracto y acabado de calidad, un proceso de cambio basado en las relaciones y acciones de una comunidad de educadores, padres de familia, políticos, administradores y personas afectadas por la educación (Wenger, 1996). El desarrollo de procesos de participación, organización de consorcios, trabajo en equipo y constante monitoría del trabajo escolar constituye la base de la creación de una comunidad nacional de aprendizaje en torno de la educación.

Valor agregado

¿Qué agrega la escuela o el programa educativo al aprendizaje de los estudiantes? Esta pregunta intenta resolver la evaluación basada en el concepto de valor agregado. Las metas establecidas y los resultados de aprendizaje de un sistema educativo o de una escuela en particular se hallan condicionados por varios factores, uno de los cuales es el rendimiento académico de los estudiantes previo a su ingreso al respectivo programa o escuela. Juzgar, por ejemplo, el trabajo de una escuela basándose exclusivamente en el puntaje obtenido por sus estudiantes en pruebas finales, sin considerar su contexto y el desempeño inicial de su población estudiantil, puede conducir a errores de interpretación sobre el trabajo de los maestros en diversas condiciones y a apreciaciones sesgadas sobre el éxito comparativo de los sistemas.

Desempeño de los docentes

Los maestros son considerados actores centrales del proceso educativo. Se supone que una vez que logren un nivel apropiado de formación o de capacitación podrán cumplir con sus funciones adecuadamente. Los bajos resultados de los indicadores de eficiencia de la educación en buena parte de los países de la región ponen en duda este supuesto. Varios países se han percatado de que es imprescindible contar con mecanismos de apoyo al desarrollo profesional de los docentes, los cuales han de incluir además, oportunidades de capacitación y sistemas de evaluación de su desempeño profesional. Los estudios y casos presentados en este libro sugieren nexos estrechos entre los desempeños de los profesores y los de sus estudiantes. La evaluación del desempeño profesional del maestro a través de la medición del rendimiento escolar de sus estudiantes no es suficiente para mejorar el sistema en su totalidad. Varios países de la región consideran que es necesario establecer criterios

y formas de evaluar también el ejercicio de la profesión docente, a pesar de que los esfuerzos del pasado han tenido poco éxito.

Existe así mismo la necesidad de hacer participar activamente a los docentes en los esfuerzos de reforma educativa y el establecimiento de estándares profesionales. Como se señaló en la discusión sobre los programas de Escuela Nueva en Colombia y las 900 Escuelas en Chile, hacer a los profesores partícipes en el diseño de las actividades de reforma puede ser exitoso en términos del mejoramiento de la calidad de la educación. Estos casos sugieren, además, que tal participación conduce a la reflexión sobre su trabajo y a una especie de autoevaluación. Estas experiencias y la aplicación de modelos como los presentados en la "Serie de Práctica" en el capítulo sobre evaluación de los maestros, sugieren la posibilidad de vencer la tradicional resistencia de los sindicatos de maestros a los esfuerzos de evaluación.

Énfasis en los consorcios

Consideraciones sobre la capacidad técnica establecida en los países, los costos involucrados en la oferta del servicio educativo y el cambio del papel de los estados, han conducido a que se aumenten los consorcios y las alianzas de diversos actores dentro y fuera del sistema educativo para llevar a cabo la evaluación educativa. En varios países los consorcios incluyen la colaboración entre ministerios de educación, universidades y centros privados de investigación. En tales casos, algunos componentes del proceso de evaluación se encargan a organizaciones dentro de la comunidad educativa pero fuera del Ministerio de Educación, las cuales pueden ejecutar ciertas tareas con mayor eficiencia.

Un segundo tipo de consorcio es la participación de un país en programas de evaluación internacional. Como se sugiere en el capítulo de evaluación del currículo en donde se describe la experiencia del proyecto TIMSS, las comparaciones del desempeño escolar entre países pueden ayudar a determinar diferencias en los objetivos, los procesos y el contenido de los programas nacionales y contribuir al desarrollo de políticas de reforma. La participación en tales programas proporciona experiencia en nuevas metodologías de evaluación que pueden aplicarse en casos nacionales. Sin embargo, los países con menores recursos deben considerar las ventajas que obtendrían de su participación, el grado de comparabilidad de sus sistemas con los de otros países y la forma de utilización de la información resultante.

Los esfuerzos regionales para desarrollar estándares comunes es el tercer tipo de consorcios descritos en este volumen. Estos consorcios ayudan a unos países a aprender de la experiencia de otros y desarrollar consensos sobre elementos críticos de cualquier reforma educativa tales como los indicadores de éxito o los mecanismos de evaluación. También permiten a los países conocer "las mejores prácticas" que pueden inspirar innovaciones en otros contextos, los tipos de trabajo estudiantil que se consideran de alta calidad o la manera como se solucionan los problemas de información. Pero estos consorcios requieren compromisos a largo

plazo y apoyo político tanto nacional como internacional.

Finalmente, los estudios presentados enfatizan el consorcio entre el gobierno—por medio de los ministerios de educación—y la sociedad civil. Esto ocurre cuando se abre el proceso de debate en torno de las reformas educativas a las personas involucradas en todos los niveles y ámbitos del sistema educativo. Aunque estos debates se han concentrado en el tema de la eficiencia del sistema más que en sus resultados en términos de aprendizaje, el problema de los estándares a los que debe responder la educación, empieza a adquirir una importancia mayor para la opinión pública en la mayoría de los países.

Capacidad local

El fortalecimiento y la utilización de la capacidad local para llevar a cabo tareas de evaluación están estrechamente relacionados con el desarrollo de intercambios institucionales. Esta capacidad está constituída por tres elementos fundamentales: una masa crítica de investigadores y analistas, instituciones eficientes y redes de información. Los datos disponibles indican que en varios países de la región existe una capacidad humana e institucional y una experiencia notable en el desarrollo y la ejecución de sistemas de evaluación del rendimiento escolar de amplia cobertura. El problema consiste en fortalecer esa capacidad, ampliar su rango de actividades de tal manera que pueda alimentar un proceso de aprendizaje social orientado a mejorar la educación y conectar la masa crítica de evaluadores e investigadores a través de redes nacionales e internacionales de información. Para algunos países con menor experiencia, se trata de crear organizaciones locales y aprender de la experiencia de otros.

Ampliación del uso de la información

Todos los trabajos presentados reflejan una preocupación por el uso de la información producida por los sistemas de evaluación. Aunque los usuarios potenciales de tal información cubren un amplio espectro de grupos, tales como los padres de familia, estudiantes, profesores, administradores, políticos e investigadores, existe sólo limitada evidencia sobre la manera como tales personas asimilan y utilizan los resultados de los sistemas de evaluación. En cierto modo aquí se presenta una situación similar a la del uso de la investigación social en general, a pesar de que los resultados de la evaluación tienden a afectar más directamente a los interesados.

Aunque los casos analizados presentan y discuten varias ocasiones en las cuales se utilizaron los resultados de las evaluaciones o estudios de forma casi lineal y directa en la toma de decisiones, el proceso por medio del cual el conocimiento permea la sociedad, se comprende y adopta, tiende a ser más amplio y menos predecible. Parece tratarse más bien de un proceso incremental y acumulativo que se sirve de múltiples canales. A medida que el interés por la educación aumenta y la sociedad civil participa más directamente en las decisiones sobre la educación, la evaluación

de sus resultados requiere de una difusión mayor y del empleo de diversos canales de información. Los estudios presentados anteriormente señalan que además de informes, la información producida por los sistemas de evaluación se comienza a distribuir a través de programas radiales, artículos periodísticos y otras estrategias.

EJECUCIÓN DE LOS SISTEMAS DE EVALUACIÓN

La comparación de los estudios de caso presentados en los capítulos anteriores proporciona una información útil para identificar las condiciones y los procesos necesarios para la organización de un sistema de evaluación educativa. Estas experiencias pueden iluminar la decisión política de los países que se encuentran dispuestos a considerar el desarrollo de programas de amplia cobertura para el seguimiento del desempeño académico de los estudiantes.

En los países que tienen una tradición más reconocida de investigación educativa y poseen una infraestructura institucional más sólida, los sistemas nacionales de evaluación tienden a ser más estables, servir a una clientela más amplia y ofrecer un potencial mayor para la utilización de la información producida, que los países con una historia más reciente en este campo. La creación de la experiencia técnica, la capacidad institucional y las redes de contactos es un proceso prolongado, pero de gran importancia para el desarrollo y uso de los sistemas de seguimiento de la educación y de sus resultados. La existencia de centros de investigación y de programas de posgrado en las universidades facilita este proceso.

Todos los países de América Latina y el Caribe poseen un sistema básico—aunque limitado—para recopilar información sobre indicadores generales de desarrollo educativo. Debido al énfasis que se puso en facilitar a un mayor número de niños el acceso a la educación, los indicadores disponibles se refieren a las tasas de matrícula, repetición y deserción escolar e incluyen alguna información sobre infraestructura física. Pocos sistemas educativos nacionales disponen de la capacidad para medir el desempeño académico más allá de las evaluaciones llevadas a cabo en el aula por los maestros.

La experiencia de los pocos países que han desarrollado sistemas nacionales de evaluación—inicialmente para propósitos de selección y posteriormente para mejorar la calidad de la educación—sugiere que existen algunos requisitos básicos para las operaciones iniciales de tales sistemas. Los principales son capacidad técnica y consenso político. La estabilidad y continuidad de tales sistemas dependen de una serie de factores entre los cuales se encuentran, además de los recursos técnicos, la capacidad de gestión, la presencia de instituciones que tengan como misión la evaluación e investigación y el desarrollo de los sistemas de información y comunicación.

La limitada experiencia de la región en programas de evaluación y monitoría internacional indica que tales programas proporcionan un impulso decisivo a programas locales, al facilitar el aprendizaje entre grupos de diferentes países sobre

la operación y eficacia de diversas alternativas. Desde este punto de vista, el papel de las agencias internacionales puede ser crítico, ya que tienen la facilidad de conectar a diferentes actores y la posibilidad de acceder a recursos no limitados para actividades locales o nacionales. Sin embargo, tales programas requieren una sólida base en los países respectivos con responsabilidades institucionales claramente establecidas, capacidad técnica y respaldo político.

OPCIONES DE POLÍTICA

Las expectativas creadas en torno de los efectos de las reformas educativas tienden a ser muy altas. Se espera que las reformas contribuyan a la solución de problemas antiguos, como la pobreza y la desigualdad y a aumentar la capacidad competitiva de las naciones. Sin embargo, en la mayoría de los países las reformas se están ejecutando en medio de ajustes económicos, escasez de recursos y disminución del tamaño de los estados. Este contexto plantea dilemas constantes políticas sociales y decisiones sobre la prioridad del gasto público. Aunque la creación y el fortalecimiento de la capacidad de juicio crítico y evaluación en el sistema en su conjunto aparecen como consecuencias lógicas de los programas de reforma, es indispensable dilucidar cuál es su contribución y cómo se garantiza su relevancia, utilidad y calidad técnica. Entre los temas de política de mayor urgencia, relacionados con las tendencias anteriormente descritas, se pueden mencionar los siguientes.

¿Debe invertirse en la evaluación educativa?

La mayor parte de los países de la región no disponen de la experiencia e infraestructura que al parecer se necesitan para la creación de sistemas de evaluación de amplia cobertura. El montaje de tales sistemas acarrea altos costos y sus resultados no son inmediatamente visibles. Una consideración inicial de política es si la inversión en la organización de los sistemas de evaluación educativa es justificable o si basta con la costumbre ya establecida de los exámenes que los profesores administran en sus respectivos cursos, para conocer el resultado de la educación e identificar a los responsables de sus éxitos o limitaciones. En este caso la pregunta sobre el logro de los "estándares nacionales" no podría responderse cabalmente, ni sería un interrogante importante. El supuesto sería que una inversión en otros rubros más directamente relacionados con la enseñanza mejoraría la educación. El argumento para apoyar esta posición sería que no necesariamente la existencia de sistemas de evaluación garantiza la mejora de la educación en un país.

Pero la opción de invertir en la creación y adecuación de mecanismos que permitan conocer la operación y los resultados de los sistemas educativos está justificada por la necesidad de facilitar la vinculación de la sociedad en su conjunto a la gestión de las escuelas, por la utilidad de la información que ilumine permanantemente las decisiones que tanto los gobiernos como los ciudadanos necesitan tomar y por la urgencia de definir nuestras propias utopías sociales y la finalidad de los sistemas educativos nacionales. Más aún, la dinámica social creada por la evaluación

constituye uno de los principales motores de las reformas puesto que, alrededor de la información que producen y de los interrogantes que despiertan, se debaten los temas críticos de la política y se esbozan nuevas direcciones.

¿Cómo debería medirse la calidad de la educación?

Como se mencionó anteriormente, qué tipo de desempeño estudiantil se mide depende en buena parte de la definición implícita o explícita que un país tenga de la calidad de la educación. Si el interés radica en las mediciones generales de la operación del sistema, los indicadores sobre escolaridad son apropiados, aunque no suficientes para conocer su rendimiento. Si se concibe la calidad en relación con los resultados del aprendizaje, un sistema de pruebas sería muy oportuno. Pruebas estandarizadas o basadas en normas es una opción que ha sido utilizada en varios países de la región. Los estudios de caso muestran, sin embargo, que existe un interés creciente por asegurarse de que todos los estudiantes logren al menos una competencia básica en lenguaje y matemáticas en la escuela primaria. Esto sugiere el desarrollo de estándares de desempeño y uso de pruebas basadas en criterios.

Como se mencionó en varios de los capítulos anteriores, ninguna de estas maneras de abordar la cuestión de la calidad de la educación responde a la concepción de la calidad como ambiente que promueve el pensamiento crítico. Una crítica común de las pruebas estandarizadas es que penalizan la imaginación y el pensamiento crítico, puesto que gratifican a los estudiantes que rápidamente seleccionan la respuesta considerada "mejor" por los diseñadores de la respectiva prueba. Mediciones complementarias, como las conocidas en la literatura bajo el nombre de "evaluación auténtica", que incluyen portafolios o archivos de los trabajos de los estudiantes, exposiciones, observaciones y otras manifiestaciones de logros de la educación, son muy oportunas. Estas técnicas pueden formar parte de los programas de capacitación de maestros en el área de evaluación del aprendizaje escolar.

"Calidad de la educación" es un concepto heurístico que probablemente evolucione con el desarrollo de nuestro conocimiento sobre el aprendizaje y la tecnología disponible. Los sistemas de evaluación facilitan esos desarrollos conceptuales y la experimentación de alternativas. Además, permiten que los logros parciales de la búsqueda permanente de mejores oportunidades de educación se viertan en la práctica a través de un proceso iterativo entre el conocimiento y la acción.

¿Qué otros aspectos de la evaluación deben considerar las políticas de la educación?

Los sistemas de evaluación existentes en la mayoría de los países se concentran en la medición del rendimiento académico. Pero, como sugieren varios de los capítulos anteriores, existen otros componentes del escenario de la educación que tienen que ser considerados en las políticas de seguimiento del sistema, tales como los procesos que conducen supuestamente a los resultados (lo que sucede en el aula, lo

que se enseña); la actuación de los profesionales responsables (maestros y directores) y la organización escolar en el centro (los ministerios y secretarías de educación) y la periferia (las escuelas).

El maestro es un actor central en el proceso de la educación. La evaluación de su desempeño profesional y su contribución al aprendizaje de los alumnos ha sido un tema controvertido. Entre los obstáculos que se han identificado para el desarrollo de procedimientos para la evaluación de los maestros, existen varios de carácter teórico: qué constituye una buena enseñanza no es algo fácil de dilucidar. Otros obstáculos son de naturaleza gremial: preocupaciones por la estabilidad laboral, o pérdida del poder de los sindicatos. Otros finalmente, son problemas de ética profesional y de justicia. Invertir en la profesionalización del magisterio o la evaluación y el seguimiento de su tarea es un dilema de política.

Dentro de un contexto de escasez de recursos, emerge otro dilema: evaluación de los estudiantes o de los maestros. A medida que la tarea de los maestros se profesionalice y que la escuela obtenga mayor autonomía, los maestros mismos estarán más involucrados en la actividad de evaluación de su trabajo. Nuevos instrumentos basados en sólidos criterios éticos, profesionales y surgidos de la investigación, están abriendo nuevos horizontes para la solución de los dilemas en medio de los cuales han oscilado las políticas existentes.

Los estudios citados en varios capítulos de este trabajo sugieren la importancia de emplear indicadores del proceso de la educación en las aulas como complemento a la evaluación de los efectos de la educación. Esta información contribuye ampliamente al debate público, al desarrollo de las políticas de currículo y a la orientación de la industria de textos y materiales educativos.

Inversión en los consorcios

Los consorcios tienen un gran potencial como apoyo a los sistemas de evaluación. En el contexto nacional, los consorcios entre el Ministerio de Educación y otras organizaciones de apoyo a la educación pueden ayudar al desarrollo y la ejecución de la evaluación de la educación de una manera eficiente, aprovechando las ventajas que cada tipo de entidad tiene. Los consorcios internacionales brindan la oportunidad de aprender sobre nuevas tecnologías y construir marcos de trabajo comunes entre diferentes naciones. Estas inversiones parecen ser muy positivas desde la perspectiva de costo-beneficio. El consorcio entre el Estado y la sociedad civil, cuando se basa en un consenso amplio sobre la necesidad de seguimiento y evaluación, puede ayudar a superar los impedimentos tradicionales a los que se haya sometida la temática, por ejemplo, de la evaluación de la profesión docente.

Los consorcios y las alianzas, tienen un costo. No se trata solamente del costo de participación, sino también de los costos implicados en la gerencia y el manejo de las relaciones entre los miembros, selección de las organizaciones y personas participantes y cumplimiento de las obligaciones emanadas de los compromisos

adquiridos. En los programas internacionales de comparación del rendimiento escolar existe el riesgo de enfatizar la competencia, descartar el análisis y disminuir el apoyo nacional a la educación frente a los resultados obtenidos, cuando éstos no están de acuerdo con las expectativas. Estos y otros temas relativos a los beneficios de participar en este tipo de consorcios, deben ser considerados de antemano en cada caso.

Inversión para fomentar el uso de la información sobre evaluación

Es importante saber qué pueden y qué no pueden hacer los sistemas de evaluación. Ellos no mejoran por sí solos la educación ni tienen la capacidad de revelar todos sus resultados. Tampoco resuelven directamente los problemas del sistema. Proporcionan, en cambio, una información muy importante a lo largo del tiempo que permite comparar, detectar problemas y ofrecer datos y análisis para las decisiones de política. Son un instrumento imprescindible para ejecutar cambios apropiados, desarrollar nuevas alternativas e involucrar a la sociedad en su conjunto en el debate sobre la educación.

La experiencia internacional en el uso de la información para el cambio social en general y el cambio de la educación en particular, ha demostrado que la información o el conocimiento, por si solos, no conducen al cambio social. Las políticas y decisiones sobre la educación no son dirigidas primordialmente ni por la investigación ni la evaluación. La formulación de la política es un proceso continuo que comprende relaciones institucionales e interpersonales, fuentes de información múltiples y procesos de negociación. La influencia de la información sobre la decisión política depende de una variedad de factores como el contexto en que se presente tal información, quién la presenta, el momento en que se presenta y la importancia y calidad de la presentación.

Existe un reconocimiento creciente de que si se quiere lograr un cambio significativo la información debe ser distribuida, debatida y analizada de tal manera que se obtenga un significado compartido entre los profesionales involucrados en los diferentes niveles del sistema educativo. Tal proceso de elaboración de conocimiento altera conceptos y puede conducir a cambios de comportamiento inspirados por tales conceptos. Esto sugiere que maestros, padres, administradores y quienes toman decisiones dentro del sistema educativo deben estar involucrados en el desarrollo, el análisis, la interpretación y la utilización del conocimiento relacionado con la enseñanza y el aprendizaje.

Una alternativa de política es invertir en la diseminación de la información. Tal diseminación no debe estar encaminada exclusivamente a comunicar a los actores decisiones tomadas, sino que debe permitirles una participación en la creación de significados compartidos, conceptos nuevos, interpretaciones e información. Esto puede llevar a una estrategia de desarrollo de los sistemas de evaluación que incluya inversiones estratégicas de involucrar a los interesados en el uso del conocimiento generado y en la producción de nuevo conocimiento sobre cómo ampliar las oportunidades de aprendizaje en la sociedad. Las posibilidades que ofrece la tecnología de

la información para el seguimiento de la educación de cada persona, del desempeño de los sistemas y de sus diferentes organizaciones y actores no tienen paralelo en la historia. El reto para la evaluación e investigación consiste en crear las condiciones para el permanente aprendizaje de las naciones.

CONCLUSIONES

Las reformas de la educación enfrentan dos problemas fundamentales. El primero es un problema filosófico: ¿Cuáles son el propósito y las características deseables de la educación en cada sociedad? ¿Cuáles son las utopías sociales que justifican una educación para todos, una escuela pública, un apoyo a la educación privada y una organización social que facilite el aprendizaje permanente? El otro es un problema tecnológico: ¿Cómo planificar y ejecutar los cambios necesarios? Los mecanismos de aprendizaje y renovación que la evaluación proporciona responden a esa doble inquietud.

Buena parte de la tecnología necesaria para la evaluación sistemática de los estudiantes, los maestros y las escuelas está disponible en los países de América Latina y el Caribe. Así mismo, la mayoría de los países disponen de una masa crítica de especialistas capaces de adelantar con éxito programas de investigación y evaluación de la educación y una limitada capacidad institucional. No obstante, la existencia de estos recursos no garantiza que se lleven a cabo evaluaciones ni que sus resultados, ni los resultados de la investigación educativa se utilicen en las decisiones de política.

Una vez que se decida invertir en evaluación educativa, existe una serie de pasos que parecen contribuir a que los resultados de las evaluaciones sean más útiles para el mejoramiento de la calidad y la eficacia de la educación en un país. Estos pasos incluyen al menos:

- El desarrollo de un consenso en torno de la agenda de evaluación e investigación que reúna las inquietudes y necesidades de una gran variedad de actores en el escenario de la educación.
- El uso del conocimiento existente por medio de la revisión de la experiencia internacional y el compromiso de las organizaciones locales que trabajan en las áreas de evaluación e investigación.
- La formación de consorcios entre el Ministerio de Educación y otras organizaciones con experiencia en investigación y evaluación con miras a realizar esfuerzos conjuntos.
- El desarrollo y la utilización de un conjunto de instrumentos analíticos para la comprensión de los fenómenos educativos con el fin de mejorar el aprendizaje de los niños.
- La utilización de los hallazgos de la evaluación para proporcionar una información de retorno que conduzca a nuevas concepciones y a acciones efectivas de tal manera que el mejoramiento continuo y el aprendizaje se conviertan en

elemento substancial de la cultura de los sistemas educativos.

NOTA

[1] Para un resumen de los estudios de la función de producción en educación, véase por ejemplo Rubin, A. (1997). "Assessing Designs for School Effectiveness Research and School Improvement in Developing Countries". *Comparative Education Review*, vol.41, 2: 178-204, Mayo.

REFERENCIAS

Gardner, Howard. (1991). "The Unschooled Mind. How Children Think and how Schools Should Teach." Nueva York: HarperCollins, Basic Books.

Heyneman, S. (1997). "Economic Growth and the International Trade in Education Reform". Trabajo presentado a la conferencia de la Agencia de los Estados Unidos para el Desarrollo Internacional, USAID. Human Capacity Development for the 21st Century, Washington, DC, julio 14-18, 1997.

Levinger, B. (1996). *Critical Transitions: Human Capacity Development across the Life Span*. Newton, MA: Education Development Center.

Wenger, E. (1996). *Communities of Practice*. Londres: Cambridge University Press.

Autores

Benjamín Álvarez H. Ph.D. Egresado de la Universidad de Nuevo México, miembro de la Academia Internacional de Educación y de la Asociación Americana de Evaluación, es Consejero de Política Educativa en la Academia para el Desarrollo Educativo en Washington, D.C.

Thomas Kellaghan es Director del Centro de Investigación Educativa de Saint Patrick's College en Dublín, Irlanda. Egresado de Queen's University de Belfast es autor y coautor de veinte libros y monografías y 120 artículos especializados. Es miembro de la Academia Internacional de Educación y de la Academia Europea

Gilbert A. Valverde es Profesor Asistente de Medición y Métodos Cuantitativos en la Universidad del Estado de Michigan y Director Asociado del Centro de los Estados Unidos que supervisa la participación del país en el Tercer Estudio Internacional de Matemáticas y Ciencias (TIMSS). Obtuvo su doctorado en la Universidad de Chicago.

María Inés Cuadros Ferré. Psicóloga de la Universidad Javeriana de Colombia. Durante los años 1994-1996 coordinó la Secretaría Pro Témpore Americana de Seguimiento a los Acuerdos sobre la Niñez, labor que realizó desde la Presidencia de la República y el Departamento Nacional de Planeación de Colombia. Trabaja actualmente para Save the Children Fund U.K., Programa para Suramérica.

Erika Himmel, especialista en Medición y Evaluación Educativa es profesora de la Universidad Católica de Chile. Dirigió el equipo que desarrolló el sistema de selección de estudiantes para las universidades chilenas y participó en la ejecución del Sistema de Medición de la Calidad de la Educación en Chile (SIMCE).

Gabriel Restrepo es profesor de Sociología de la Universidad Nacional de Colombia. Ha sido Presidente de la Asociación Colombiana de Sociología y Jefe de la Unidad de Desarrollo Social del Departamento Nacional de Planeación.

Carol Anne Dwyer es la Directora Ejecutiva de la División de Investigación de Programas y Políticas Educativas del Servicio de Pruebas Educativas en los Estados Unidos (Educational Testing Service, ETS). Es la autora principal de la serie de publicaciones sobre la evaluación de los maestros que inician su carrera profesional, desarrollada por el Servicio de Pruebas Educativas (ETS).

Francisco Álvarez Martín es investigador del Centro para la Investigación y el Desarrollo de la Educación (CIDE) de Chile. Es Licenciado en Pedagogía y experto en capacitación de docentes. **Paula Vergara y María José Álvarez** participan también en el equipo de investigación del CIDE.

William J. Webster es Director de la División de Sistemas de Investigación, Evaluación e Información de las escuelas públicas de Dallas. Obtuvo su doctorado en la Universidad del Estado de Michigan. Ha publicado más de 50 artículos especializados en evaluación y estadística aplicada.

Robert L. Mendro, Ph.D. de la Universidad de Colorado, dirige la Unidad de Investigación de las escuelas públicas de Dallas.

María Teresa de la Fuente es consultora en el proyecto de mejoramiento de la calidad de la educación en el Estado de Paraná.

Heloísa Luck, doctora en educación de la Universidad de Columbia, es consultora en la Secretaría de Educación de Paraná.

Corina Lucía Ramos, obtuvo su maestría en educación en la Universidad Federal de Paraná. Actualmente es consultora en planificación y evaluación de la gestión educativa.

Margarita Zorrilla es Directora de Desarrollo Educativo del Instituto de Educación de Aguascalientes. Es también profesora de la Universidad Autónoma de Aguascalientes.

Ray Chesterfield, Vice Presidente de Juarez and Associates, obtuvo su doctorado en educación internacional y antropología. Su trabajo se ha dirigido a la creación y evaluación de programas que promueven la formación de los recursos humanos de comunidades marginadas.

BIBLIOGRAFÍA RECIENTE

Teresa Kavanaugh

Airasian, Peter. 1993. "Policy-Driven Assessment Of Assessment Driven-Policy?" Measurement and Evaluation in Counseling and Development, vol. 26.

La finalidad general de las evaluaciones impulsadas por la política consiste en producir un cambio educativo y mejorar el desempeño de los alumnos, maestros y el sistema educativo en general. Más concretamente, dos miras guían la evaluación impulsada por la política. Primero, las evaluaciones se utilizan como mecanismo para mejorar el desempeño de los maestros y los alumnos poniendo de relieve estándares y motivando a los alumnos y maestros a trabajar con más ahínco. Segundo, las evaluaciones se utilizan para obtener control sobre los objetivos de la educación y, en algunos casos, el proceso de enseñar, a fin de concentrarse, aclarar e influir en lo que se enseñó en las aulas y las escuelas. Declara que las evaluaciones impulsadas por la política tienen un efecto simbólico importante. Las pruebas y evaluaciones simbolizan el orden, el control y los resultados deseables de las escuelas y son símbolos morales poderosos de una serie tradicional de valores sociales y educativos como el trabajo duro y la recompensa por el esfuerzo. El artículo declara que un nombre más exacto para lo que se ha llevado a cabo en nombre de la reforma educativa en la década precedente es política impulsada por las evaluaciones, no evaluación impulsada por la política. Llega a la conclusión de que existe una tendencia discernible y un conflicto probable que guardan relación con la política futura de evaluación. La tendencia entraña esfuerzos por ampliar la evaluación de la dependencia principal en formatos de preguntas de opciones múltiples a una mayor dependencia en la evaluación auténtica. El conflicto incluye al nivel local, estatal o nacional en el que la mayoría de las evaluaciones influyentes se llevarán a cabo en la próxima década.

Banco Mundial. 1995. Priorities and Strategies for Education: A World Bank Review. Washington, DC: Banco Mundial.

El capítulo seis de este libro analiza la importancia de los resultados educativos en el establecimiento de prioridades y estrategias para la educación. Argumenta que, hasta la fecha, se ha prestado una atención insuficiente a los resultados educativos bien se determinen por el mercado laboral o en términos de aprendizaje. Trata del

uso de los resultados para establecer y vigilar las prioridades públicas, con énfasis especial en el análisis de la tasa de rendimiento. Aborda la cuestión del establecimiento de estándares y la vigilancia del desempeño en la educación, temas que entran en juego cuando el sector público en un determinado país ha adoptado decisiones relacionadas con la asignación de los recursos públicos. El libro indica que, una vez que se han establecido estándares de desempeño, es necesario observar este último. Especifica que la forma más común de hacerlo es mediante pruebas y exámenes. Finalmente, el capítulo termina con un esbozo de las aplicaciones normativas y pedagógicas de los exámenes y las pruebas públicas.

Capper, Joanne. 1996. Testing to Learn/ Learning to Test. Washington, DC: Academia para el Desarrollo Educativo

Con demasiada frecuencia, las pruebas educativas y evaluaciones nacionales miden el aprendizaje superficial (es decir, la memorización de hechos) en vez de la comprensión y dominio de los conceptos. Este libro está concebido como guía integral para mejorar las pruebas de la educación en los países en desarrollo. Aborda la relación entre los exámenes y las evaluaciones, y la enseñanza y el aprendizaje en estos países. Si bien los capítulos dos al siete se concentran en los aspectos más técnicos del diseño y la ejecución de las pruebas, los capítulos uno y ocho proporcionan un conocimiento valioso de las cuestiones relacionadas con las pruebas y las evaluaciones al nivel nacional.

Departamento de Evaluación de las Operaciones, Banco Mundial. 1994. "Building Evaluation Capacity." Lessons and Practice. Http:// www.worldbank.org/html/oed/lp004.htm#evaluation. Washington, DC: Banco Mundial.

Esta ponencia examina los pasos requeridos para establecer y beneficiarse de la capacidad de evaluación del sector público en los países en desarrollo. Esboza formas en las que la evaluación puede desempeñar un papel vital en cuatro áreas de la gestión del sector público de un país: (1) influir en el análisis y la formulación de política, (2) mejorar la asignación de los recursos y el proceso presupuestario, (3) mejorar los programas y proyectos de inversión, (4) examinar misiones fundamentales de las instituciones o el propio gobierno. Esboza los problemas actuales con la evaluación en los países en desarrollo. Muchos países en desarrollo aún carecen de los requisitos esenciales para una evaluación eficaz: la calidad de la información y el acceso a ella son a menudo deficientes, los mecanismos de reacción al proceso de toma de decisiones son débiles y la cultura de responsabilidad no está firmemente instituida. El informe proporciona lecciones para los planificadores, subrayando que la evaluación puede desarrollarse óptimamente si todas las partes interesadas–dentro del país y en la comunidad en desarrollo–la consideran como una forma de aprender y mejorar el desempeño del sector público. Analiza cómo formular una estrategia específica para el país y proporciona observaciones sobre el desarrollo de la capacidad de evaluación.

Greany, Vincent, y Kellaghan, Thomas. 1995. "Equity Issues in Public Examinations in Developing Countries." Documento Técnico del Banco Mundial No. 272, *Asia Technical Series*, Washington, DC: Banco Mundial.

Este estudio presenta un análisis de las inequidades asociadas con los exámenes públicos en los países en desarrollo. Se analiza la investigación procedente de cerca de treinta países, principalmente en África y Asia. Debido a la importancia que se da al desempeño en los exámenes, los maestros enseñan el examen y, como resultado, las oportunidades para los alumnos que salen de la escuela a una edad temprana son inadecuadas. Se presentan las prácticas asociadas con los exámenes que pueden crear inequidades. Entre éstas figuran procedimientos de puntuación, el uso de preguntas culturalmente inapropiadas, el requisito de que los candidatos paguen honorarios, el uso de tutores privados, el examen en un idioma con el que algunos candidatos no están familiarizados y toda una serie de prácticas inapropiadas. El informe advierte que el uso de los sistemas de cuotas para hacer frente a las diferencias en el desempeño asociadas con la ubicación, la etnicidad o un grupo lingüístico también crea desigualdades para algunos alumnos. Llega a la conclusión de que la evidencia limitada disponible no indica que los exámenes creen desigualdades entre los géneros y que la clasificación de las escuelas sobre la base del desempeño de los exámenes de los estudiantes puede no proporcionar una evaluación justa del trabajo de las escuelas.

Greany, Vincent, y Kellaghan, Thomas. 1996. Monitoring the Learning Outcomes of Education Systems. Washington, DC: Banco Mundial.

Por muchos años, los gobiernos han recopilado y publicado datos estadísticos sobre cómo sus sistemas de educación están funcionando y desarrollándose, en particular concentrándose en indicadores tales como números e instalaciones de escuelas, matrículas de estudiantes e índices de eficiencia tales como las relaciones de estudiante-maestro y las tasas de repetición, deserción y conclusión de la cohorte. Sin embargo, muchos países han recopilado y brindado sistemáticamente información sobre los resultados de la educación. Esta obra se concentra en la vigilancia de los resultados de aprendizaje de los sistemas educativos como medio de mejorar la calidad del aprendizaje de los alumnos. Proporciona un análisis general de la naturaleza y opciones de los indicadores del resultado, un estudio de las evaluaciones nacionales e internacionales, evaluaciones nacionales y exámenes públicos, componentes sugeridos para una buena evaluación nacional basada en los resultados y, estudio de caso ficticio de una evaluación nacional con el objetivo de analizar los lineamientos previamente esbozados en el libro. El libro distingue el uso de los indicadores de los resultados como los que miden el desarrollo cognoscitivo y afectivo de los alumnos. Estudia las diferencias, ventajas y desventajas entre las evaluaciones nacionales e internacionales. También examina las diferencias entre las evaluaciones nacionales y los exámenes públicos. Finalmente, ofrece lineamientos a las personas interesadas en garantizar una evaluación nacional eficaz, afirmando que no sólo los interesados deberían estar de acuerdo sino que la

parte interesada más importante e influyente la constituye el Ministerio de Educación.

Greany, Vincent y Kellaghan, Thomas. (s.f.)"The Integrity of Public Examinations in Developing Countries." Mimeo.

Este documento se concentra en examinar cómo se observan o infringen los procedimientos en virtud de los cuales se prepararan, administran y evalúan los exámenes. El informe subraya que sólo si se aplican con éxito dichos procedimientos habrá seguridad de que se mantiene la integridad, solidez y uniformidad de los exámenes de forma que ningún candidato tenga una ventaja o desventaja en relación con otros candidatos debido a una práctica desleal y que, en consecuencia, las notas o grados otorgados a los candidatos están directamente relacionados con la capacidad que se mide en vez de con factores irrelevantes o condiciones no controladas. El estudio analiza las fuentes de evidencia disponibles sobre el tema de la práctica indebida y las formas de prácticas indebidas[1] que se han identificado en los exámenes. Considera las razones para las prácticas indebidas y esboza algunos de los procedimientos que se han elaborado para detectarlas. Finalmente, el informe explora formas de controlar o evitar su incidencia.

Horn, Robin; Wolff, Laurence; y Vélez, Eduardo. 1991. "Establecimiento de Sistemas de Medición del Rendimiento Académico en América Latina: Un Análisis de los Problemas y la Experiencia Más Reciente." América Latina y El Caribe, Departamento Técnico, Documento del Banco Mundial No. 9. Washington, DC: Banco Mundial.

Esta obra esboza formas de diseñar y administrar pruebas y analiza factores importantes que deberían tomarse en consideración en este proceso. Se citan las experiencias de Chile, Costa Rica, México y Colombia como experiencias eficaces en la evaluación nacional. Sin embargo, aunque estos países han tenido las evaluaciones nacionales más eficaces, ninguno de ellos han incorporado componentes que estén dedicados a la investigación y al desarrollo de las propias pruebas. El informe subraya el hecho de que, para garantizar que las pruebas logren realmente una mayor calidad en la educación, los países deberían tomar en cuenta la importancia de definir los objetivos de largo plazo de las metas de las evaluaciones así como un plan para la divulgación de los resultados de las pruebas.

Kellaghan, Thomas, y Greany, Vincent. 1996. "Using Examinations to Improve Education: A Study in Fourteen African Countries." Documento Técnico del Banco Mundial No. 165. Africa Technical Department Series. Washington, DC: Banco Mundial.

Este informe resumido presenta una descripción de las clases, funciones, niveles de desempeño, dirección, administración y financiamiento de los exámenes públicos basados en una serie de estudios emprendidos sobre los exámenes de primaria y

secundaria en catorce países africanos. Se esbozan procedimientos para financiamiento de los exámenes; construcción, administración y puntuación de los trabajos y declaración de los resultados en cada país. Si bien la función principal de los exámenes públicos en estos países es elevar los estándares académicos y seleccionar a los alumnos para el siguiente nivel más alto del sistema educativo, las conclusiones de este informe indican que, si bien estos exámenes pueden ayudar a elevar los estándares académicos, también pueden contribuir a problemas en el sistema educativo.

Madaus, George y Kellaghan, Thomas. 1991. "Student Examination Systems in the European Community: Lessons for the United States." Informe por Contrato presentado a la Oficina de Evaluación de Tecnología. Washington, DC: Congreso de los Estados Unidos.

En la primera parte de este informe, se examina la evolución de la política de pruebas y se consideran seis propuestas para establecer exámenes nacionales en los Estados Unidos. La segunda parte analiza los orígenes y el desarrollo de los exámenes públicos en Europa. Describe la complejidad, operación y contextos de los sistemas de examen de estos países, concentrándose en particular en Francia, Alemania y el Reino Unido. Esboza las tres características principales de estos sistemas de exámenes públicos: su función selectiva, la parte importante que desempeñan las universidades y el papel de los exámenes en definir el aprendizaje de los estudiantes–qué aprenden los estudiantes y cómo lo aprenden. El documento considera las repercusiones de estos exámenes para la escolaridad en los Estados Unidos.

Mislevy, Robert. 1994. "What Can we Learn from International Assessments?" Ponencia presentada en la Conferencia sobre el Uso de los Datos Educativos Internacionales, Washington, DC.

Se ha considerado que las evaluaciones internacionales producen información que permite comparaciones de resultados relativos por país y materia o que permite el mejoramiento en un país a partir de las determinantes de los resultados en otros, o finalmente como forma de proporcionar información a los órganos normativos sobre el estado de desempeño y prácticas en sus propios países. En esta ponencia, se exploran las clases de inferencias que pueden sacarse de la evaluación educativa internacional, considerando la evidencia que puede obtenerse y cómo puede interpretarse. Se argumenta que los índices de los resultados educativos que son comparables en grados distintos entre países pueden ser útiles pero que hay que reconocer que la determinación de la situación relativa de los países proporcionará muy poca información acerca de cómo obtener la política educativa o mejorar la práctica de la instrucción.

Murphy, Paud; Greany, Vincent; y Lockheed, Marlaine. 1996. National Assessment: Testing the System. Washington, DC: Banco Mundial (Economic Development Institute).

El presente libro está concebido para proporcionar a los responsables de la política y los profesionales de la educación que trabajan en los países en desarrollo información acerca de aspectos diferentes de los sistemas nacionales de evaluación y la forma en que funcionan en la práctica. Esboza cierto número de características diferentes de las evaluaciones nacionales y las personas interesadas en el estudio de dichos sistemas. Está concebido para promover un debate de la calidad de la provisión de la educación en los países, cómo debería medirse y cómo utilizar la información recopilada para mejorar la calidad. La obra está escrita principalmente para órganos normativos y profesionales en los ministerios de educación. Sin embargo, también tienen por fin beneficiar a quienes tratan de obtener información acerca de distintos aspectos técnicos de las evaluaciones nacionales y cómo funcionan estos sistemas en los países en desarrollo.

O'Neil, John. 1993. "Can National Standards Make a Difference?" Educational Leadership, vol. 50, no. 5.

Este artículo analiza y examina las cuestiones y el progreso hasta la fecha en el movimiento de estándares nacionales en los Estados Unidos y termina esbozando algunas preocupaciones suscitadas por quienes critican los estándares nacionales. Cita la obra del Consejo Nacional sobre Estándares de Educación y Pruebas como clave para aclarar metas y definiciones de estándares nacionales. El Consejo pidió un sistema de educación basado en estándares supervisado por un sistema nacional de exámenes de los alumnos. Concretamente, el Consejo recomendó que: deberían elaborarse estándares nacionales que incluyan estándares de contenido (qué deberían conocer los estudiantes y qué deberían poder hacer), estándares del desempeño de los alumnos (el nivel o niveles de competencia de los alumnos en el contenido) y estándares de desempeño del sistema (para evaluar el éxito de las escuelas, distritos, estados y el país en su conjunto en ayudar a todos los estudiantes a alcanzar altas normas de desempeño). Además, el Consejo declaró que los estados deberían formular estándares de provisión de servicios de las escuelas para enjuiciar el hecho de si las escuelas están proporcionando a los alumnos la oportunidad de alcanzar estos estándares elevados. El Consejo también declaró que debería establecerse un sistema de evaluación nacional vinculado con estos estándares.

Ravitch, Diane. 1997. National Standards in American Education: A Citizen's Guide. Washington, DC: Brookings Press.

Los defensores de los estándares nacionales declaran que estándares claros y uniformes mejorarían los resultados académicos y prepararían a los alumnos del país para hacer frente a los retos del siglo XXI. Los detractores consideran que los estándares serían demasiado homogéneos e inflexibles para la población cultural y económicamente diversa y que darían demasiado poder al gobierno federal para dictaminar lo que debería enseñarse a los niños. Esta obra explora la promesa de un sistema nacional de estándares y los problemas que rodean su ejecución. Pone de relieve las experiencias de los Estados Unidos y saca comparaciones con otros países. Resume el caso a favor y contra un sistema nacional de estándares y

evaluaciones, expresado en el contexto del debate en los Estados Unidos. Después describe el debate: Por qué aumentó el consenso a favor de la reforma, cómo aumentó el ímpetu a favor del cambio y el marco contra el que siguió adelante la cuestión.

En este estudio, el término práctica indebida se utiliza para describir cualquier acción prescrita adoptada en relación con un examen o prueba que tiene como intención obtener una ventaja desleal o, en algunos casos, colocar al candidato en desventaja.

Resnick, Lauren, y Nolan, Kate. 1995. "Where in the World are World Class Standards?" Educational Leadership, vol. 52, no. 6.

Los países conocidos por sus alumnos extraordinarios tienen varias prácticas en común; a la cabeza de la lista figuran estándares de educación pública claros, uniformes y exigentes. En un esfuerzo por elaborar una definición de los estándares de clase mundial, el Proyecto de Nuevos Estándares de la Universidad de Pittsburgh inició una actividad internacional de establecimiento de referencia en 1993 con la esperanza de recopilar y analizar documentos normalizados de otros países. En un intento por obtener un cuadro claro del desempeño en los estándares de clase mundial, los autores examinan el seguimiento, currículo y exámenes en los Estados Unidos, los Países Bajos, Francia, Suecia y Alemania. El proyecto determinó que las prácticas compartidas y los elementos comunes entre enfoques muy diferentes para la educación en nuestros países enseñan importantes lecciones: el establecimiento de estándares claros, uniformes, exigentes y públicos ayuda a los alumnos a obtener buenos resultados; el seguimiento y las prácticas de agrupación han de ser lógicos en la cultura de la escuela y en las metas futuras del alumno y de la comunidad; los exámenes deberán someter a prueba lo que se ha pedido a los estudiantes que aprendan, preferiblemente de la misma forma que han de desempeñarse en la clase; los exámenes que requieren tareas complejas y exigentes pueden administrarse a una amplia gama de estudiantes, quizás a todos ellos. Como profesionales de primera línea en el proceso educativo, los maestros deberían tener gran poder decisorio sobre lo que constituye los exámenes y la forma en que se evalúan.

Sanders, William, y Horn, Sandra. 1995. "Educational Assessment Reassessed: The Usefulness of Standardized and Alternative Measures of Student Achievement as Indicators for the Assessment of Educational Outcomes." Educational Policy Analysis Archives, Vol. 3, No. 6

Los métodos de evaluación basados en el uso de pruebas estandarizadas han sido sometidos a ataques intensos en años recientes y algunos detractores han llegado hasta pedir su total eliminación. El presente documento esboza el debate de si la evaluación estandarizada o la evaluación de alternativa es un modelo mejor para evaluar los resultados de la educación. Indica que la determinación de los indicadores educativos más apropiados para evaluar si los estudiantes han logrado

las metas establecidas para ellos es un elemento central del debate. El documento expone que las pruebas estandarizadas proporcionan indicadores viables, poco costosos, fiables y válidos del aprendizaje de los estudiantes que son particularmente útiles en la evaluación de las entidades educativas y el desempeño de los alumnos. También subraya la importancia de otras posibles formas de evaluación como instrumentos viables para evaluar el progreso y resultado de los alumnos en tanto se presta una especial atención a garantizar su validez y fiabilidad. El documento concluye defendiendo el uso de indicadores múltiples del aprendizaje de los alumnos, incluidos los proporcionados por las pruebas estandarizadas.

Schiefelbein, Ernesto. 1993. "The Use of National Assessments to Improve Primary Education in Chile." En: From Data to Action: Information Systems in Educational Planning, editado por David W. Chapman y Lars O. Mahlick, Ed. París: UNESCO.

Aunque en los últimos 20 años se han emprendido en Chile tres evaluaciones nacionales del desempeño de los alumnos como base para mejorar la calidad de la educación, estas actividades no han generado los incrementos previstos en los resultados de los alumnos y han aumentado la igualdad entre los alumnos procedentes de diferentes grupos socioeconómicos. Esta ponencia describe las miras, el diseño, la operación y el impacto de los tres programas. Una conclusión importante fue la extraordinaria estabilidad en las puntuaciones del desempeño de los estudiantes de cuarto grado en el curso de este período de tiempo, interpretados por algunos como prueba de estancamiento del sistema educativo. Este estudio de caso explora posibles causas de dicha estabilidad, sugiere mejoras en el uso de los resultados de las pruebas y examina el grado en que los análisis utilizados para efectuar comparaciones entre grupos de escuelas y entre distintos años son apropiados.

Snyder, Conrad. 1997. "Exam Fervor or Fever: Case Studies of the Influence of Primary Leaving Examinations on Uganda Classrooms, Teachers, and Pupils." Washington, DC: Academia para el Desarrollo Educativo.

En muchos sistemas de educación, los exámenes definen el éxito y el fracaso de los individuos, actúan como medios de acceso a oportunidades futuras y proporcionan credibilidad a los sistemas que los utilizan. Los Exámenes de Salida de Primaria (PLE) de Uganda ilustran estas características. Este informe aborda varias cuestiones clave que exploran la influencia de los PLE en las aulas, maestros y padres de Uganda. Dada la influencia general de los exámenes, ¿cómo afectan a los niños y adultos que participan en ellos? ¿Puede influir la naturaleza específica de los exámenes en la forma en que se imparte la educación? ¿Son estas estrategias simples que pueden utilizarse para manipular el enfoque del examen de forma que surta un efecto concentrado en el aula? El informe declara que las respuestas a estas preguntas son: (1) los exámenes en Uganda surten efectos sustanciales en los

individuos afectados, (2) los exámenes sí influyen en la forma en que se imparte la educación, (3) el impacto de los exámenes no es simple. El informe concluye indicando que los exámenes nacionales son componentes complicados de un sistema complejo. Aunque las evaluaciones nacionales proporcionan influencia para el control normativo central sobre la instrucción en el aula, su uso en la reforma no es tan simple como pudiera parecer a primera vista.