



Establecimiento de un marco para la evaluación e incentivos docentes

CONSIDERACIONES PARA MÉXICO



Establecimiento de un marco para la evaluación e incentivos docentes

CONSIDERACIONES PARA MÉXICO

Este trabajo se publica bajo la responsabilidad del Secretario General de la OCDE. Las opiniones expresadas y las conclusiones formuladas en este informe no necesariamente corresponden a las de los gobiernos de los países miembros de la OCDE.

Por favor, cite esta publicación de la siguiente manera:

OCDE (2011), *Establecimiento de un marco para la evaluación e incentivos docentes: Consideraciones para México*, OECD Publishing. <http://dx.doi.org/10.1787/9789264059986-es>

ISBN 978-92-64-05999-3 (impreso)

ISBN 978-92-64-05998-6 (PDF)

Esta publicación es producto del acuerdo de cooperación celebrado entre el Gobierno de los Estados Unidos Mexicanos y la Organización para la Cooperación y el Desarrollo Económicos, que busca mejorar la calidad de la educación en México.

Photo credit: ©UNESCO/José Gabriel Ruiz Lembo.

Las erratas de las publicaciones de la OCDE se encuentran en línea en www.oecd.org/publishing/corrigenda.

© OCDE 2011

Usted puede copiar, descargar o imprimir los contenidos de la OCDE para su propio uso y puede incluir extractos de publicaciones, bases de datos y productos de multimedia de la OCDE en sus propios documentos, presentaciones, blogs, sitios web y materiales docentes, siempre y cuando se dé el adecuado reconocimiento a la OCDE como fuente y propietaria del copyright. Toda solicitud para uso público o comercial y derechos de traducción deberá dirigirse a rights@oecd.org. Las solicitudes de permisos para fotocopiar parte de este material con fines comerciales o de uso público deben dirigirse al Copyright Clearance Center (CCC), info@copyright.com o al Centre français d'exploitation du droit de copie (CFC), contact@cfcopies.com.

Prefacio

La educación es la base para un futuro exitoso de nuestras sociedades. Del mismo modo, los docentes son la base del éxito del sistema educativo de una nación. De hecho, un sistema efectivo combina diversos y numerosos elementos, como son los contenidos curriculares y los estándares nacionales, el liderazgo escolar y el desempeño de las escuelas, la calidad, la motivación y las perspectivas de los docentes, así como un sistema eficaz de evaluación educativa. Sin embargo, los docentes son indispensables, por lo que muchos gobiernos ponen ahora una mayor atención al papel que juegan.

Este informe, *Establecimiento de un marco para la evaluación e incentivos docentes: Consideraciones para México*, examina tanto el diseño como la implementación gradual de políticas eficaces en torno a la evaluación e incentivos docentes. Junto con sus publicaciones hermanas, *Mejorar las escuelas: Estrategias para la acción en México* (2010) y *Evaluación y reconocimiento de la calidad de los docentes: Prácticas internacionales* (2009), este libro presenta una estrategia amplia de políticas públicas para el proceso de reforma educativa en México.

Ciertamente, México ha registrado un avance impresionante en la matrícula estudiantil, y más recientemente, en la construcción de un sólido marco institucional para evaluar los resultados del aprendizaje de sus alumnos. Pero aún quedan cosas por hacer.

En primer lugar, México debe mejorar todavía su planta docente. Para que su desempeño educativo progrese de “adecuado” a “bueno” y después de “bueno” a “excelente”, México deberá conseguir que el prestigio de la profesión docente iguale el de otras profesiones más valoradas. Un paquete completo de reformas para atraer a los mejores graduados a la profesión docente y para convertirlos en instructores eficaces, requerirá mejorar las prácticas pedagógicas mediante el uso de mejores prácticas de capacitación y contratación, mediante la reforma del sistema de compensaciones y salarios, y poniendo en marcha incentivos adecuados y diferenciados.

En segundo lugar, México debe concentrarse en las “tres Es” e implantar un sistema de Evaluación Educativa más Eficaz. Es necesario medir el éxito de los esfuerzos educativos a partir de los resultados de aprendizaje de los alumnos. Al mismo tiempo que se fortalezcan el sistema y las herramientas actuales, será indispensable contar con más herramientas de medición y evaluación. Es necesario que estas herramientas se basen cada vez más en el desempeño, vinculen mejor la información entre la enseñanza y los resultados del aprendizaje, y formen parte de un sistema de enseñanza-aprendizaje completo y bien alineado.

Aconsejamos al gobierno mexicano a emprender las reformas necesarias mediante: *i)* el fortalecimiento de su sistema de evaluación centrado en los resultados del aprendizaje de los alumnos; *ii)* el fortalecimiento de sus políticas docentes, tales como tomar las medidas necesarias hacia la evaluación docente; y *iii)* asegurar que todos los actores estén comprometidos y motivados para mejorar el desempeño. La OCDE se declara lista para acompañar a México en la puesta en marcha de su agenda amplia de reformas educativas.

La elaboración de este informe contó con la orientación del *Consejo Asesor de la OCDE sobre Políticas de Evaluación e Incentivos Docentes en México*, un conjunto de expertos de renombre internacional y analistas de la OCDE, encabezados por Carlos Mancera. Los siguientes miembros del Consejo Asesor y expertos invitados contribuyeron a este informe: José Luis Gaviria, Jorge Juárez Barba, Enrique Roca Cobo, Halsey Rogers, Lucrecia Santibáñez, Susan Sclafani, Margarita Zorrilla, Leonel Zúñiga, Mathew Springer y Eva Baker.

El informe se publica bajo el patrocinio de la División de Indicadores y Análisis del Directorado de Educación de la OCDE, bajo la responsabilidad de Alejandro Gómez Palma, Marlene Gras, Andreas Schleicher, Michael Davidson, William Thorn, Elisabeth Villoutreix, Isabelle Moulherat, Diana Tramontano, Iván Salinas, Laura Milena Valencia y Marika Boiron.



Angel Gurría,
Secretario General de la OCDE

Agradecimientos especiales

Desde el inicio en 2008 del Acuerdo de Cooperación para mejorar la calidad de la educación, la OCDE y el gobierno de México han colaborado de manera estrecha para apoyar las reformas educativas actuales y para las del futuro. El trabajo ha centrado su atención en: la eficiencia en el diseño e implementación de políticas públicas, la medición y la evaluación, los incentivos, las políticas docentes y el desempeño escolar. Este texto presenta las conclusiones y principales consideraciones para apoyar las reformas en dichas áreas. Como parte de esta labor, la OCDE dispuso de sus conocimientos y su experiencia, revisó las prácticas internacionales relevantes y participó en consultas con expertos nacionales e internacionales, así como con actores relevantes en México, mediante la realización de reuniones técnicas, seminarios internacionales y visitas de observación.

El *Consejo Asesor de la OCDE sobre Políticas de Evaluación e Incentivos Docentes en México* agradece profundamente al Secretario de Educación de México Alonso Lujambio y a los numerosos departamentos de la Secretaría de Educación Pública de México por su presteza para compartir información, por su disposición para compartir su saber, así como su interés en los resultados de nuestros esfuerzos conjuntos. En particular, queremos agradecer a Francisco Ciscomani, encargado de la Unidad de Planeación y Evaluación de Políticas Educativas (UPEPE), y a su equipo, sobre todo Bernardo Rojas, Lourdes Saavedra, Silvia Ojeda y Janina Cuevas, cuyo profesionalismo y cuya calidad humana hicieron de este desafiante reto una tarea aun más grata. Asimismo, deseamos aprovechar esta oportunidad para agradecer a los funcionarios de alto nivel que apoyaron nuestra labor: Fernando González, subsecretario de Educación Básica, y Rodolfo Tuirán, subsecretario de Educación Superior. Muchos directores generales de la Secretaría contribuyeron también al éxito de nuestros esfuerzos conjuntos: Ana María Aceves, Rafael Freyre Martínez, Marcela Santillán, Leticia Gutiérrez y Juan Martín Martínez. Por su tiempo, apoyo y dedicada contribución quedamos profundamente agradecidos. Silvia Ortega, rectora de la Universidad Pedagógica Nacional, proporcionó ideas muy pertinentes a lo largo del proceso. Asimismo, nos gustaría extender nuestra gratitud a Josefina Vázquez Mota, ex secretaria de Educación, a Jorge Santibáñez, ex encargado de la UPEPE, y a sus respectivos equipos por su interés y apoyo, en particular en las primeras etapas del Acuerdo de Cooperación entre la OCDE y México.

Quedamos en deuda con todas las autoridades estatales educativas mexicanas que apoyaron nuestros esfuerzos y realizaron contribuciones valiosas, en especial las de Aguascalientes, Chiapas, Nuevo León y Veracruz. Nuestra labor, y las recomendaciones resultantes, se beneficiaron de su generoso apoyo e ideas. El interés y compromiso de diversos expertos nacionales e internacionales, organizaciones de la sociedad civil, legisladores y representantes del Sindicato Nacional de Trabajadores de la Educación (SNTE) son asimismo motivo de un gran agradecimiento. El éxito de las reformas en México, como en toda nación, dependerá en gran medida de la pluralidad y visión común que compartan los actores relevantes.

Asimismo queremos agradecer a todos los alumnos, directores escolares, maestros, supervisores, jefes de zona y a los Asesores Técnico Pedagógicos, quienes compartieron su experiencia, retos, preocupaciones y esperanzas con nosotros. Su intervención representa un saludable recordatorio de que el éxito final de las reformas se medirá en sus salones de clases y escuelas.

Deseamos extender nuestro reconocimiento al embajador Agustín García López, representante permanente de México ante la OCDE, y también a Luisa Solchaga, consejera de Educación, y a su equipo, por su apoyo y contribuciones. El *Consejo Asesor de la OCDE sobre Políticas de Evaluación e Incentivos Docentes en México* expresa un agradecimiento especial a su presidente, Carlos Mancera, por su dedicado compromiso y valiosa aportación a lo largo del proceso.

Para consultar información sobre este y otros informes relevantes producidos por el Acuerdo OCDE-México para Mejorar la Calidad de la Educación en Escuelas en México, favor de visitar la página electrónica: www.oecd.org/edu/calidadeducativa.

Índice

RESUMEN EJECUTIVO.....	9
CAPÍTULO 1 MÉXICO RESPONDE A LOS DESAFÍOS EDUCATIVOS.....	19
CAPÍTULO 2 MARCO DE POLÍTICAS PÚBLICAS PARA IMPLEMENTAR LAS REFORMAS EDUCATIVAS.....	27
2.1 Reformas de políticas educativas en un contexto internacional.....	28
2.2 Movilización de la investigación de la OCDE, de las prácticas internacionales y del conocimiento nacional.....	30
2.3 Ejes de política pública para las reformas educativas: La importancia de articular las preguntas correctas.....	33
2.4 Consideraciones para México.....	34
CAPÍTULO 3 LA RENDICIÓN DE CUENTAS COMO IMPULSOR DE POLÍTICAS PARA MEJORAR LOS RESULTADOS DEL APRENDIZAJE.....	39
3.1 Tipos y características de los sistemas de rendición de cuentas en la educación.....	40
3.2 Desempeño docente: Hacia una comprensión más amplia de la rendición de cuentas.....	42
3.3 Consideraciones para México.....	44
CAPÍTULO 4 USAR LOS RESULTADOS DEL APRENDIZAJE DEL ALUMNO PARA MEDIR LAS MEJORAS.....	49
4.1 Resultados del aprendizaje del alumno: Instrumentos y medidas de evaluación.....	51
4.2 El sistema de evaluación ENLACE en México.....	56
4.3 Desafíos y oportunidades para fortalecer el sistema de evaluación ENLACE.....	61
4.4 Resumen de recomendaciones para México.....	63
CAPÍTULO 5 EL VALOR AGREGADO DE LAS ESCUELAS: MEDICIONES MÁS JUSTAS Y EQUITATIVAS.....	69
5.1 Los modelos de valor agregado y la escuela como unidad de rendición de cuentas.....	70
5.2 Importancia de la calidad de los datos y la información.....	75
5.3 Consecuencias vinculadas a una evaluación más justa y fiable de escuelas y maestros.....	76
5.4 Consideraciones para México.....	76
CAPÍTULO 6 LA EVALUACIÓN DE DOCENTES EN SERVICIO: ASPECTOS DE LAS POLÍTICAS Y DE LA IMPLEMENTACIÓN.....	81
6.1 Prácticas internacionales.....	82
6.2 Cuatro preguntas clave que deben abordar los sistemas de evaluación.....	84
6.3 Consideraciones para México.....	85
CAPÍTULO 7 LOS INCENTIVOS PARA DOCENTES EN SERVICIO.....	93
7.1 Tipos de incentivos docentes.....	95
7.2 Directrices nacionales e implementación local: Importancia del equilibrio adecuado.....	106
7.3 Programas piloto, monitoreo y evaluación de incentivos.....	107
7.4 Consideraciones para México.....	108
Apéndice 7A Las pruebas piloto, el monitoreo y la evaluación de incentivos docentes: Prácticas recomendadas.....	111
CONCLUSIÓN.....	127

Recuadros

Recuadro 4.1	<i>Prova Brasil</i> para la rendición de cuentas y las mejoras.....	51
Recuadro 4.2	Sistemas mixtos de evaluaciones de alumnos.....	55
Recuadro 5.1	Jurisdicciones en México.....	71
Recuadro 5.2	Ejemplo de un modelo de valor agregado de regresión lineal.....	73
Recuadro 7A.1	Definiciones importantes.....	111
Recuadro 7A.2	Importancia de la participación de los actores educativos.....	119
Recuadro 7A.3	La Campaña de Calidad de Datos.....	120

Gráficas

Gráfica 2.1	Representación esquemática del proceso de movilización, análisis y aplicación eficaces del conocimiento (Heurística específica para un país).....	32
Gráfica 2.2	Representación esquemática de un modelo heurístico específico al contexto local de país: Marco de Políticas Públicas para Implementar Reformas Educativas.....	34
Gráfica 3.1	Representación esquemática del concepto del umbral de capacidad en la rendición de cuentas docente.....	43
Gráfica 5.1	Representación esquemática de un modelo sencillo de valor agregado.....	72
Gráfica 7.1	Relación entre la prueba piloto, el monitoreo y la evaluación de un programa.....	112

Cuadros

Cuadro 1.1	Áreas cubiertas por la <i>Alianza por la Calidad de la Educación</i> y el Acuerdo OCDE-México.....	20
Cuadro 1.2	Metodología de la OCDE para apoyar la implementación de políticas educativas y resultados del Acuerdo de Cooperación.....	20
Cuadro 1.3	Dimensiones del sistema de educación básica, año lectivo 2008-2009 (formal).....	22
Cuadro 1.4	Educación básica por tipo de escuela, año lectivo 2008-2009.....	22
Cuadro 3.1	Tipos de evaluación más comunes usados en sistemas de rendición de cuentas.....	42
Cuadro 4.1	Instrumentos y fuentes de evidencia para evaluar el aprendizaje del alumno.....	52
Cuadro 4.2	Fiabilidad de ENLACE.....	59
Cuadro 4.3	Correlación entre sub-escalas de resolución de problemas, lectura y ciencias, PISA 2003.....	60
Cuadro 4.4	Porcentaje de trampa probable detectado en ENLACE de 2006 a 2009.....	61
Cuadro 5.1	Beneficios e implicaciones de los métodos de valor agregado para la rendición de cuentas y la mejora escolares.....	77
Cuadro 6.1	Panorama general de las prácticas de evaluación docente en México.....	86
Cuadro 6.2	Implementación de pasos en orden secuencial.....	86
Cuadro 6.3	Resumen de recomendaciones específicas para México sobre la evaluación docente.....	87
Cuadro 7.1	Resumen de los beneficios de realizar programas piloto.....	108
Cuadro 7A.1	Enfoque y propósito de las preguntas de evaluación en el contexto de programas de incentivos.....	113
Cuadro 7A.2	Distintos diseños de evaluación para investigar el impacto del programa y las intervenciones de las políticas educativas.....	114
Cuadro 7A.3	Enfoque y propósito de las preguntas de evaluación en el contexto de incentivos.....	119

Resumen ejecutivo

OPORTUNIDAD PARA REFORMAS EDUCATIVAS EN MÉXICO

México, la décimo cuarta economía más grande del mundo (2009), enfrenta importantes desafíos en la educación. Pese al progreso significativo alcanzado en las décadas recientes en términos del acceso, de las mejoras en los índices de terminación en los niveles de educación básica, y del desarrollo de evaluaciones del aprendizaje, aún persisten desafíos considerables. México invierte ya un alto porcentaje del presupuesto público en educación (cerca del 22%, el más alto entre los países de la OCDE). En este contexto mejorar la calidad de los servicios educativos, incrementar los niveles de logro académico y reducir las tasas de deserción son temas prioritarios. A la par, es de igual importancia que México se asegure de que todos los niños y jóvenes, incluyendo aquellos provenientes de entornos socioeconómicos desfavorecidos y de familias indígenas, tengan las mismas oportunidades educativas.

Con el propósito de abordar estos temas, el gobierno mexicano estableció prioridades para las reformas educativas en su Programa Sectorial de Educación 2007-2012. Para dar seguimiento al avance de los objetivos previstos hasta su consecución, la Secretaría de Educación Pública (SEP) estableció indicadores de mejora para el aprendizaje de los alumnos tal como son medidos por la Evaluación Nacional del Logro Académico en Centros Escolares (ENLACE) y el Programa para la Evaluación Internacional de Alumnos de la OCDE (PISA, por sus siglas en inglés). Otros indicadores clave se relacionan con el desarrollo profesional de los maestros, la toma de decisiones en el ámbito escolar, la equidad en las oportunidades educativas y las reformas relacionadas con el contenido y los planes de estudio. Para impulsar el proceso de reforma educativa, el gobierno mexicano estableció en 2008 la *Alianza por la Calidad de la Educación* con el Sindicato Nacional de Trabajadores de la Educación (SNTE).

En este contexto, la SEP y la OCDE establecieron en 2008 el *Acuerdo para Mejorar la Calidad de la Educación de las Escuelas en México*. El propósito del acuerdo fue determinar no sólo qué cambios de política había que hacer en México, sino también cómo debían diseñarse e implementarse eficazmente las reformas de la política educativa, tomando en cuenta las iniciativas en marcha, así como las condiciones, restricciones y oportunidades locales. Uno de los componentes de este acuerdo implicó el desarrollo de políticas y prácticas adecuadas para evaluar la calidad de las escuelas y los maestros y para vincular los resultados con los incentivos con el fin de generar procesos efectivos de mejora. Estas líneas de trabajo fueron dirigidas por el Consejo Asesor de la OCDE sobre Políticas de Evaluación e Incentivos Docentes en México, conformado por expertos internacionales.

Los principales resultados del acuerdo de cooperación entre la SEP y la OCDE se presentan en esta y en las siguientes publicaciones: *Mejorar las escuelas: Estrategias para la acción en México* (OCDE, 2010) (elaborado por el Consejo Asesor de la OCDE sobre Gestión Escolar y Política Docente en México), y *Evaluación y reconocimiento de la calidad de los docentes: Prácticas internacionales*, editado por Susan Sclafani para la OCDE (2009). Como parte de este esfuerzo, también se actualizó y tradujo material sobre modelos de valor agregado: *La medición del aprendizaje de los alumnos: Mejores prácticas para evaluar el valor agregado de las escuelas* (OCDE, 2011). A continuación se presenta un resumen de las principales recomendaciones, enfocadas en los temas de evaluación e incentivos para maestros, desarrolladas por el Consejo Asesor y el Secretariado de la OCDE durante el Acuerdo de Cooperación.

1. El marco de políticas públicas para implementar las reformas educativas

Para que las autoridades educativas y actores clave adapten y pongan en marcha reformas de política pública diseñadas a partir de las mejores prácticas y ejemplos internacionales, deberán considerarse de manera adecuada las condiciones, restricciones y oportunidades locales. Al combinarse con prácticas internacionales y evidencia comparable, la movilización del conocimiento local puede proporcionar un vínculo vital para diseñar reformas educativas adecuadas a las prioridades y el contexto nacionales. El propósito de las recomendaciones en este ámbito es proporcionar lineamientos a la SEP, y a los actores implicados en México, para una continua movilización del conocimiento local, con el fin de informar los procesos de la reforma educativa.

1.1 En combinación con las prácticas internacionales y la evidencia disponible, la **movilización del conocimiento específico del país** es un elemento vital para diseñar, planificar y poner en marcha reformas educativas de manera eficaz que sean viables y sostenibles dadas las condiciones, restricciones y oportunidades que hay en México. Las experiencias sugieren con claridad que el conocimiento confiable y actualizado sobre temas particulares de política pública es crucial durante el proceso de adoptar mejores prácticas y recomendaciones de otros contextos.

1.2 *Existen dimensiones de política comunes que deberían considerarse para programas de reforma exitosos.*

Las dimensiones de la política pública proporcionan un marco mediante el cual la SEP, y los actores implicados, pueden continuar el proceso de análisis, diseño, planificación, implementación y evaluación de políticas educativas. Los programas de reforma educativa en México se beneficiarían de la consideración metódica de cada una de las siguientes **seis dimensiones del marco de política pública**:

- **Calidad de datos, información e indicadores:** esto implica una consideración de la calidad y la cantidad de los datos pertinentes y la información disponible (sobre alumnos y maestros, desempeño y vínculos entre ellos), para fijar los objetivos y la identificación de áreas deficientes que necesiten mejorarse.
- **Relevancia social del compromiso de los actores implicados:** esto incluye considerar opciones para la estrategia de comunicación, participación y consulta con los actores implicados, incluidos el público en general, los maestros, los directores y las autoridades educativas locales, así como los sindicatos docentes (lo cual implica identificar de qué manera la reforma propuesta puede traducirse en un mensaje socialmente pertinente y significativo para todas las familias, los maestros y los directores).
- **Financiamiento público adecuado:** es importante considerar la cantidad y la permanencia del financiamiento público para el desarrollo e implementación de la reforma educativa (el cual puede ser, por ejemplo, anual o fijo), incluyendo un análisis del costo-beneficio, las proyecciones de costo y los ahorros que se pueden obtener al modificar la asignación actual de partidas presupuestales o programas existentes.
- **Marco legal y normativo adecuado:** es importante identificar conflictos potenciales y posibles modificaciones que pudieran requerirse en áreas relacionadas (por ejemplo, en las leyes laborales) para llevar a cabo las reformas.
- **Arreglo institucional:** lo cual incluye un análisis de las instituciones públicas (SEP, INEE y autoridades educativas estatales, por ejemplo, para identificar organismos específicos que deberían contribuir al desarrollo de estándares y evaluaciones, y para proponer modificaciones).
- **Procesos de descentralización y transferencia de responsabilidades:** esto significa prestar atención a los niveles *formales* y *de facto* de la transferencia de responsabilidades en los principales organismos federales y estatales responsables de proporcionar servicios educativos (incluyendo los recursos, la capacidad, la gestión de información, la evaluación y la supervisión).

2. La rendición de cuentas pública

Todos los actores implicados deberán sentirse responsables y tendrán que rendir cuentas públicamente del desempeño de los alumnos y de los resultados educativos en general.

2.1 El desempeño, la equidad y una mayor eficacia de la inversión en educación son desafíos para México, como lo son para otros países que emprenden reformas educativas importantes. Esto se plantea por el resultado de México en las evaluaciones internacionales (como la de PISA), por la gran diversidad que existe entre y al interior de las regiones y de los estados mexicanos, así como por la importancia que la inversión en la educación sigue teniendo en términos de proporción del gasto público, a pesar de su gasto moderado por alumno comparándolo con el de otros países de la OCDE. Involucrar a todos los actores del sistema educativo en México en la rendición de cuentas para mejorar los resultados de todos los alumnos, en todas las escuelas, transmite un mensaje claro y constituye una manera de alinear esfuerzos y recursos.

2.2 Los actores deben rendir cuentas acerca del avance en el aprendizaje de los alumnos, y se les debe proporcionar la asistencia técnica y la formación de capacidades necesarias. Un sistema de rendición de cuentas en México, definido con claridad y enfocado en los resultados de aprendizaje y crecimiento de los alumnos, puede aportar la alineación necesaria de distintos esfuerzos dados el tamaño, la complejidad y los múltiples intereses de los participantes del sector educativo. Utilizar el aprendizaje de los alumnos como un criterio clave, del que deberán rendir cuentas las autoridades educativas estatales, las escuelas, los directores y los maestros, refleja las tendencias internacionales que se enfocan en los resultados y se alejan de las reformas orientadas únicamente en los insumos. Las prácticas internacionales relativas a los incentivos para los maestros basados en el desempeño reflejan este cambio. No obstante, esto no significa que los temas relacionados con la infraestructura o la inclusión social ya no sean importantes para el sector educativo mexicano, más bien, que el aprendizaje y el desarrollo para todos los alumnos – medidos y evaluados desde diferentes puntos de vista – deberán ser el objetivo final de las reformas de política educativa. Del mismo modo, el apoyo a los alumnos, las escuelas, los directores y los maestros, así como el desarrollo profesional, constituyen elementos vitales para incrementar la rendición de cuentas.¹

2.3 La rendición de cuentas centrada en el aprendizaje y el logro de los alumnos implica establecer estándares claros de contenido y de desempeño. El desarrollo de estándares como componente clave del sistema de rendición de cuentas centrado en el aprendizaje de los alumnos debería enfocarse por lo menos en tres prioridades: *i*) el desarrollo apropiado de estándares de contenido y estándares de desempeño de alumnos y maestros; *ii*) alineación curricular y coherencia entre los estándares, la evaluación y el desarrollo profesional; y *iii*) correspondencia de los estándares con las mejores prácticas internacionales y con los referentes internacionales del conocimiento y las habilidades de los alumnos. Dentro de un marco de rendición de cuentas basado en estándares, los actores deberán tener incentivos para cumplir, o rebasar incluso, las expectativas sintetizadas en los estándares.

2.4 Las medidas de rendición de cuentas deberían incluir criterios complementarios relacionados con el esfuerzo, así como con el desempeño. En México, un sistema de rendición de cuentas basado en estándares debe apoyarse en los resultados del aprendizaje y desarrollo de los alumnos (desde evaluaciones estandarizadas y otros métodos confiables, siempre que sea posible), así como en otros criterios complementarios sobre el desempeño individual, grupal y de la escuela. Esto es pertinente para México debido a que la asistencia, la puntualidad y el tiempo efectivo de enseñanza siguen siendo asuntos de primer orden. La rendición de cuentas en México debe considerar que algunos directores y maestros tal vez no se desempeñen de acuerdo con sus capacidades actuales. Por tal motivo, es necesario crear incentivos para aumentar el esfuerzo y el desempeño básicos, así como para apoyar la formación de capacidades y el desarrollo profesional. Por ejemplo, la reducción eficaz de los índices de deserción o la eficiencia terminal de los alumnos en una escuela pueden ser también consideradas como referentes. De igual forma, la rendición de cuentas implica que se pueda aconsejar a los maestros que reciben una ayuda técnica adecuada y oportunidades para el desarrollo profesional, y que son incapaces de mejorar su desempeño, que dejen la profesión.

- 2.5 Debe ponerse el acento en los alumnos, las escuelas y los maestros para conseguir una mejora continua tomando a la escuela como una unidad básica para la rendición de cuentas.**² Si bien diferentes niveles y actores del sistema educativo deberán rendir cuentas, la escuela puede servir como unidad básica para la rendición de cuentas, con datos individuales, información y supervisión de los alumnos y los maestros. Los datos y la información sobre los alumnos y los maestros en el ámbito escolar pueden utilizarse para concebir iniciativas específicas, tales como los programas de mejora, los modelos de crecimiento, los incentivos y estímulos para los maestros, intervenciones educativas para aquellos con bajo desempeño, así como la identificación de buenas prácticas para configurar y desarrollar estándares de enseñanza.
- 2.6 Es importante definir un proceso gradual para desarrollar un marco de evaluación que utilice fuentes múltiples de evidencia.** Establecer un sistema de rendición de cuentas basado en estándares es un proceso que se elabora en forma gradual, con etapas claras y con enfoques complementarios de medición y evaluación. Tanto las evaluaciones sumativa y formativa del aprendizaje de los alumnos, como el desempeño de la escuela y el desempeño de los maestros, deben formar parte del sistema de rendición de cuentas en México. Sin embargo, la elaboración de un sistema como éste debe planearse en todas sus etapas, poniendo una especial atención en las capacidades, los métodos y los costos que se tienen actualmente y los que se esperan alcanzar.

3. La importancia de los resultados del aprendizaje de los alumnos

El avance en el aprendizaje de los alumnos debería ser un criterio clave para medir el desempeño de las escuelas, los maestros, los organismos de participación de los padres de familia, las instituciones educativas estatales y federales, y el sistema en su conjunto. Los resultados de las evaluaciones estandarizadas son un medio relevante, pero deberán emplearse otras medidas válidas y confiables del aprendizaje de los alumnos para obtener una idea más completa de su logro.

- 3.1 El aprendizaje y logro de los alumnos como base de la rendición de cuentas requiere una serie de mediciones múltiples, con referencias recíprocas, válidas y confiables.** Como toda medición, las medidas e instrumentos de aprendizaje y logro de los alumnos (evaluaciones de los maestros, portafolios con trabajos de los alumnos, pruebas estandarizadas y observación en el aula, entre otros), presentan fuentes potenciales de errores y sesgos, por lo que se deberá crear un método en distintas etapas que use evidencia válida sobre el aprendizaje de los alumnos y su crecimiento a través del tiempo, para que tome en cuenta las herramientas que existen actualmente en México, que estime costos y determine la formación de capacidades, y el desarrollo de instrumentos que se requieren a futuro. Con estándares claros de contenido y desempeño de lo que se espera que los alumnos sepan y sepan hacer, podrán desarrollarse más y mejores medidas y procedimientos para evaluar el aprendizaje y el avance de los alumnos.
- 3.2 El uso de datos sobre el desempeño de los alumnos debe acompañarse, siempre y cuando sea posible, por medidas complementarias y confiables del aprendizaje de los alumnos, conforme estas se desarrollen, prueben y validen.** La importancia relativa de los datos de los alumnos y las evaluaciones basadas en la escuela, o por ejemplo aquellas realizadas por los maestros, pueden redefinirse según lo requieran los objetivos de política y las consecuencias derivadas de las evaluaciones. Los sistemas educativos de Australia, Hong Kong-China y Canadá, son ejemplos de sistemas que intentan combinar evaluaciones estandarizadas con aquellas basadas en la escuela (por ejemplo, las que son calificadas en el ámbito local pero moderadas en lo externo), con proyectos de alumnos y ensayos extendidos.
- 3.3 Los datos sobre el desempeño de los alumnos, como los de ENLACE, pueden desempeñar en México una función importante en los programas de rendición de cuentas y mejora escolar.** Los programas actuales de la SEP, y otras autoridades educativas estatales, demuestran el alto grado de aceptación social y potencial de ENLACE. Los datos de desempeño de los alumnos agregados por escuela, zona escolar o entidad federativa, pueden ser utilizados en modelos analíticos estáticos o de crecimiento, dependiendo del propósito específico de los programas y políticas.

- 3.4 La validez y la confiabilidad demostradas por ENLACE brindan a México una valiosa oportunidad** de explotar el potencial de los datos sobre desempeño de los alumnos proporcionados de manera anual por la evaluación. Los datos acerca de los alumnos y un modelo apropiado pueden ser utilizados para conseguir una mayor rendición de cuentas por parte de las escuelas y maestros (colectivamente), pero también para reforzar la mejora, las intervenciones educativas y la canalización de los recursos adicionales para los alumnos con bajo desempeño.
- 3.5 Debería establecerse un programa para fortalecer la evaluación ENLACE tomando en cuenta los temas de la demanda cognitiva, el alineamiento curricular y la coherencia.** Convendría considerar la mejor evidencia disponible sobre el desarrollo del aprendizaje de los alumnos. El desarrollo de ENLACE debe fijar etapas y objetivos claros que aborden consideraciones técnicas (*i.e.* la equiparación vertical), administrativas (*i.e.* identificadores únicos para los alumnos, maestros y escuelas, e interrelaciones) y logísticas (mejor supervisión de las pruebas).³ Con los usos incrementados de ENLACE en el futuro, se deberá considerar una mayor supervisión y seguridad en la administración de las pruebas. El programa debería tener también una visión de largo plazo que tome en cuenta referencias internacionales y metas de desempeño desarrolladas de manera consecuente. Durante todo el proceso se debe tener en mente la alineación y la coherencia entre los estándares, la evaluación y el desarrollo profesional de los maestros. Una visión clara del marco de evaluación en México debe explicar los distintos aunque complementarios propósitos de las diferentes evaluaciones (es decir, ENLACE, EXCALE o posibles evaluaciones basadas en la escuela) y de qué forma deberían continuar desarrollándose en el futuro dentro de un marco nacional común.
- 3.6 Con datos sobre el desempeño de los alumnos y modelos de análisis apropiados, se pueden identificar aquellas escuelas con bajo desempeño, con alto desempeño y los casos que necesitan mayor observación para darles seguimiento.** Conforme el proceso de medición y evaluación se vaya estableciendo, se pueden vincular algunas consecuencias, como son los incentivos, una observación adicional y asistencia técnica para escuelas y maestros. Esto implica un desarrollo gradual del proceso y la posibilidad de que haya consecuencias y respuestas por etapas a los resultados de las evaluaciones.

4. Evaluación equitativa del valor agregado de las escuelas

Todos los alumnos, sin importar cuáles sean sus antecedentes socioeconómicos, étnicos o lingüísticos, deberán tener las mismas oportunidades de aprender y lograr niveles altos de desempeño. Si bien se ha demostrado que el desempeño de los alumnos está bastante relacionado con los antecedentes familiares, los resultados de las mediciones y las evaluaciones deberán reflejar la verdadera contribución de la enseñanza al aprendizaje de los alumnos, y no sólo el contexto socioeconómico de la escuela o de sus alumnos.

- 4.1 Dada la gran diversidad de los contextos educativos entre y al interior de las entidades federativas mexicanas, los modelos de valor agregado pueden ofrecer una medida equitativa y más precisa del avance de los alumnos y del desempeño de la escuela.**⁴ Los esfuerzos actuales de la SEP y de las autoridades educativas estatales, en cuanto a la presentación y el uso de los resultados de ENLACE, son un buen punto de arranque. Los desafíos técnicos implicados en el diseño, la planificación y la implementación de un sistema de evaluación que utilice modelos de valor agregado, deberán abordarse con rigor durante todas las etapas. La robustez y la frecuencia de la evaluación ENLACE en México ofrecen, en este sentido, una valiosa oportunidad.
- 4.2 Los modelos de valor agregado pueden ofrecer una mejor opción que las calificaciones brutas de las pruebas para identificar de manera precisa y equitativa la contribución de las escuelas al aprendizaje de los alumnos.** Mediante el mismo proceso de evaluación y análisis, los alumnos, las escuelas y los maestros con un desempeño más bajo que el esperado, podrán ser identificados y contar con la oportunidad de mejorar mediante la ayuda de recursos adicionales y asistencia técnica.

4.3 Dadas las condiciones actuales del sistema educativo en México, los modelos de valor agregado se pueden basar tomando a la escuela como la unidad de rendición de cuentas, aunque las zonas escolares,⁵ los grupos de alumnos, los municipios y los estados también se pueden utilizar para el análisis y el diseño de acciones educativas. La equiparación vertical debe ser uno de los primeros asuntos técnicos que hay que revisar al momento de enfrentar un desarrollo adicional de ENLACE. También deben evaluarse la calidad y la disponibilidad de la información que pueda usarse para diseñar modelos *contextualizados* de valor agregado.

4.4 Las primeras etapas del desarrollo en México de un proceso de reformas, basado en modelos de valor agregado, pueden concentrarse en ejercicios analíticos realizados por las autoridades educativas (sin consecuencias), utilizando los resultados reales de los alumnos para identificar los puntos fuertes y débiles de los diferentes modelos de valor agregado. Aun antes de aplicar los métodos de valor agregado a los datos, las escuelas podrían agruparse dependiendo de los contextos socioeconómicos y se podrían utilizar modelos de logro contextualizado como posibles precursores de análisis de valor agregado.

Es por esto que el proceso de establecer un modelo de valor agregado puede tener diferentes etapas:

- *Estratificación* de escuelas similares (basadas en el tipo de escuela y en información socioeconómica u otra información pertinente) para comparar los resultados brutos entre escuelas dentro del mismo grupo. También deberán identificarse y abordarse los asuntos relacionados con la calidad de la información contextual disponible.
- *Ejercicios internos* analíticos basados en distintos modelos de valor agregado, realizados por las autoridades educativas, para escoger los modelos y resolver asuntos técnicos con datos. Cuando los datos lo permitan, se sugiere utilizar para su diseño **una media móvil de tres años**. Además, las autoridades educativas podrían utilizar el análisis para supervisar y evaluar políticas, programas y jurisdicciones específicos, como es el caso del *Programa Escuelas de Calidad*, poniendo un énfasis en las diferencias entre y al interior de municipios, zonas escolares, estados y grupos étnicos.
- *Estrategia de comunicación y consulta con actores clave* sobre los méritos, los desafíos y las oportunidades de los modelos de valor agregado. Esto podrá vincularse con un relanzamiento público de ENLACE, por ejemplo, con un programa claro para su futuro desarrollo.
- *Atribuir consecuencias* (de bajo impacto al principio) para las escuelas de bajo desempeño (exploración, observación y asistencia adicionales), así como para aquellas con alto desempeño. Los mismos análisis de valor agregado podrían ser utilizados por la SEP y las autoridades educativas estatales para identificar a las escuelas que puedan tener maestros y prácticas que valga la pena emular y promover.

5. Evaluación de los maestros para la rendición de cuentas y la mejora

Los maestros son vitales para el aprendizaje de los alumnos. Sin embargo, es difícil mejorar lo que no se mide. Por tanto, un proceso equitativo y confiable de evaluación de los maestros en servicio debería permitir que los maestros de todos los niveles del espectro de desempeño puedan mejorar, ser reconocidos y contribuir a los resultados educativos generales.

5.1 Deben desarrollarse estándares destinados a los maestros para proporcionarles no sólo una guía clara sobre qué se considera una buena práctica docente, sino también oportunidades de desarrollo y mejora profesionales. Los ejemplos y modelos de estándares internacionales pueden brindar un punto de referencia muy útil para adaptar y desarrollar más tarde estándares de enseñanza apropiados para México,⁶ los cuales deben cumplir con los siguientes criterios:

- cubrir todos los dominios de enseñanza definidos;
- establecer diferentes niveles de competencia para cada aspecto específico que defina las áreas de trabajo de los maestros y las escuelas;
- definir un grupo básico de características de rendimiento que deberán cumplir todos los maestros en todas las escuelas;

- definir y hacer operativos los objetivos y resultados propuestos de la buena enseñanza; y
- ser dinámicos para asegurar un escalamiento apropiado.

Los estándares también deben cubrir al menos las siguientes áreas: uso del tiempo de la enseñanza (asistencia, puntualidad, tiempo efectivo de clase), planificación y preparación (el diseño de las actividades de enseñanza y los procedimientos de evaluación para todos los alumnos), ambiente en el aula (hacer del aula un sitio seguro para fomentar la curiosidad intelectual), la enseñanza (atractiva y desafiante, adaptada a diferentes alumnos), y responsabilidades profesionales. Debe prestarse una especial atención a la capacidad de los maestros para velar por la equidad: atender las necesidades propias a la diversidad de los alumnos y lograr resultados adecuados de aprendizaje para todos.

5.2 Es vital el consenso entre los actores implicados sobre la importancia de desarrollar un marco de evaluación de los maestros en servicio amplio, transparente y equitativo. Dada la importancia que tienen los maestros para el aprendizaje de los alumnos, es necesario diseñar y planificar un marco de evaluación para los maestros en servicio en el corto, mediano y largo plazos. Se debe considerar la formación gradual de capacidades en diferentes niveles para asegurar la equidad y objetividad del marco de evaluación, por ejemplo, incluyendo un grupo de evaluadores externos bien capacitados. En este contexto, es importante contar con una *estrategia de comunicación, compromiso y consulta* con los principales actores implicados (incluyendo a la ciudadanía, líderes de opinión y maestros). Otros aspectos importantes de la política pública, como son las implicaciones legales, normativas y financieras (sugeridas por el marco de políticas públicas presentado en el capítulo 2), pueden asegurar que se están estableciendo bases sólidas de manera transparente y participativa, incluso si el sistema completo de evaluación docente no se implementa durante el mandato de una sola administración gubernamental.

5.3 En el contexto de un incremento de la rendición de cuentas, y junto con la posibilidad de crear capacidades y generar un desarrollo profesional de los maestros, es importante asegurarse que los maestros cumplan con los niveles mínimos de desempeño profesional y de resultados. El crecimiento en el aprendizaje de los alumnos debería formar parte de los criterios de evaluación. Sin embargo, en las primeras etapas del marco de evaluación docente también se pueden incluir aspectos básicos, tales como la asistencia, la puntualidad y el tiempo efectivo dedicado a la enseñanza-aprendizaje como medio para que todos los maestros aumenten su desempeño. La inclusión de criterios básicos como éstos puede producir ganancias considerables y oportunas para el sistema de evaluación de una manera eficiente en costos (*i.e.* asegurándose que se coseche primero la fruta que “está al alcance de la mano”).

6. Incentivos y estímulos para los maestros en servicio

Si bien las recompensas al desempeño se han utilizado de manera eficaz en otras áreas de trabajo, su uso reciente en el sector educativo, en particular a los maestros, aún está en proceso de exploración y evaluación. Por ende, la SEP, las autoridades educativas estatales y los actores implicados necesitarán determinar la combinación específica de incentivos y estímulos más eficaces, tanto monetarios como no monetarios, para México. No obstante, independientemente de las recompensas o consecuencias vinculadas con los resultados, para que un maestro sea considerado como eficaz, sus alumnos deberán demostrar niveles satisfactorios de crecimiento del aprendizaje; del mismo modo, ningún maestro deberá ser calificado como ineficaz si sus alumnos muestran niveles satisfactorios de crecimiento.

6.1 Para conseguir una política eficaz y sostenida de incentivos para los maestros en servicio, los siguientes cinco principios deberán guiar su desarrollo:

- Los incentivos deben reflejar la calidad de la enseñanza. El criterio para decir que un programa de incentivos tiene éxito no sería tanto una mejor remuneración a los maestros que se desempeñan mejor, sino el que los maestros contribuyan en mejorar los resultados del aprendizaje de sus alumnos.

- El sistema de incentivos debe reconocer y apoyar al maestro individual, al equipo de maestros de la escuela y a la profesión en su conjunto. Los incentivos deberán estar integrados en un sistema que apoye la mejora continua tanto de alumnos, maestros y escuelas, como la del sistema educativo. En el largo plazo, los incentivos basados en pruebas deberán complementarse por una sólida capacidad de manejo de los recursos humanos en las escuelas y en los ámbitos locales para evaluar con precisión la calidad del trabajo, utilizando fuentes externas de validación y corroboración pertenecientes tanto al gobierno como a la profesión docente.
- El sistema de incentivos deberá basarse en una comprensión sólida de lo que motiva a los maestros, y abarcar múltiples dimensiones de motivación, con el propósito de fomentar un ambiente de trabajo atractivo, crear y facilitar el avance en la trayectoria profesional, ofrecer acceso al desarrollo profesional, e identificar y promover prácticas eficaces de enseñanza. Por ende, los incentivos y estímulos deben considerar incentivos financieros y no pecuniarios, tales como las condiciones laborales, material suficiente para escuelas y salones de clase, reconocimiento social, mejores oportunidades de formación y desarrollo profesional, o una combinación de lo anterior.
- El sistema de incentivos debe proporcionar buenos mecanismos de retroalimentación y acceso al desarrollo profesional para garantizar así que los maestros que no reciban los incentivos comprendan qué pueden hacer para mejorar su desempeño y tener incentivos para cambiar su conducta. También deberá fomentar una cultura basada en la evidencia y en los datos.
- El sistema de incentivos deberá recompensar tanto el buen desempeño como la mejora relativa, y debe considerar el valor agregado de maestros y escuelas independientemente de los factores socioeconómicos. Sin embargo, aunque se estén elaborando modelos de valor agregado, se pueden emplear métodos más sencillos para constatar que los alumnos, escuelas y maestros están siendo comparados con otros en contextos similares (*i.e.* modelos de estratificación socioeconómica y/o de logro contextualizado).

6.2 Es importante distinguir con claridad las políticas de incentivos para maestros en servicio de otros programas relacionados con la enseñanza que, siendo semejantes, de manera básica no les ofrecen a los maestros incentivos para que mejoren su desempeño. Las políticas de incentivos deben ser comunicadas a los profesionales de la enseñanza de forma clara antes de informarles sobre las evaluaciones y mediciones que se emplearán para otorgar las recompensas y premios. Además, cada maestro elegible deberá tener la probabilidad de ser recompensado por un desempeño sobresaliente mayor a cero, lo cual no sucede en la actualidad con los programas similares existentes en México. Es por esto que será de particular importancia encontrar un equilibrio entre las directrices nacionales para los incentivos (y un financiamiento parcial por lo menos), y la flexibilidad y la coparticipación a nivel estatal por parte de las autoridades educativas locales en los recursos (financieros o de otro tipo) para incentivos y estímulos para los maestros. Por último, se recomienda realizar una prueba piloto de los posibles programas de incentivos para asegurar la viabilidad y la eficiencia en términos de costo-beneficio de su diseño. Los programas piloto deberán ser vigilados y evaluados, con líneas base claras, y con tanto control como sea posible, para ser más útiles y valiosos (por ejemplo, pruebas aleatorias o cuasi-experimentales, si las condiciones lo permiten).

6.3 Los incentivos para los maestros en servicio en México deben motivarlos para mejorar su desempeño, sin descuidar que la escuela sea la unidad básica de rendición de cuentas, tomando en cuenta el estado actual y las perspectivas de los sistemas de datos, así como la calidad de la información disponible. No obstante, aun con sistemas de datos e información que no son lo bastante sólidos, las autoridades educativas locales pueden desarrollar medidas para confirmar y validar la elegibilidad de los maestros para los incentivos (por ejemplo, con inspecciones de campo y la validación de datos sobre los alumnos, la escuela y los docentes). A medida que se desarrolle un sistema de evaluación del maestro individual sólido y creíble, las políticas de incentivos podrían modificarse para permitir que cada docente pueda recibir incentivos de forma individual en el futuro.⁷ Para los incentivos que se proporcionan de manera colectiva a las escuelas,

éstas deberán rendir cuentas públicas por los recursos adicionales recibidos. Si las escuelas pueden decidir la asignación de los recursos provenientes de los incentivos, los mecanismos deben asegurar la transparencia y la participación progresiva de los actores implicados relevantes, como es el caso de los padres de familia y los consejos escolares locales.

6.4 Los incentivos y estímulos, financieros y no financieros para los maestros, deben basarse en un proceso de medición y evaluación justo y adecuado. Dada la gran diversidad del sistema educativo mexicano, es necesario desarrollar un proceso de evaluación válido y confiable para identificar a los maestros que califican para recibir los incentivos. El éxito de los incentivos está vinculado en forma directa con la credibilidad y la equidad del proceso de medición y evaluación en el cual se basan. Se deben tomar en cuenta modelos que contemplen la diversidad socioeconómica de los alumnos mexicanos, así como otros factores que pueden influir en gran medida en su desempeño, como es el español como segunda lengua o el origen étnico, al momento de hacer comparaciones entre las escuelas y sus maestros. Las escuelas y programas de educación especial, así como las escuelas de nivel preescolar, deberían ser evaluados e incentivados mediante mediciones apropiadas del desempeño y el aprendizaje del alumno – siempre que esto sea posible. Dada la diversidad existente entre y al interior de los estados, las políticas públicas de incentivos también deben considerar un bono adicional relativo para las escuelas rurales desfavorecidas, como compensación respecto a otras escuelas urbanas o rurales no desfavorecidas. Del mismo modo, los incentivos deben apoyar la mejora continua de las escuelas y los maestros en todo el espectro del desempeño.

NOTAS

1. En el informe *Mejorar las escuelas: Estrategias para la acción en México* de la OCDE (2010) se proporcionan recomendaciones concernientes al apoyo, la formación de capacidades y el desarrollo profesional de los docentes.
2. La unidad de rendición de cuentas (*accountability* en inglés) se refiere al nivel con base en el que se supervisan, miden y evalúan el esfuerzo, las capacidades y el desempeño de los alumnos, docentes y directores. Si bien los alumnos son evaluados de manera individual, por ejemplo, y los docentes deberían estar motivados individualmente en mejorar sus prácticas, al final los resultados deben agruparse de tal modo que los individuos se sientan responsables de manera colectiva a nivel de la escuela.
3. En el capítulo 5 se encuentran las recomendaciones específicas sobre los aspectos técnicos, administrativos y logísticos que deberían considerarse en un programa para mejorar la evaluación ENLACE.
4. Un análisis detallado de los beneficios, características y posibles diseños de modelos de valor agregado se encuentran en la publicación actualizada *La medición del aprendizaje de los alumnos: Mejores prácticas para evaluar el valor agregado de las escuelas* (OCDE, 2011).
5. En México, la zona escolar es la designación administrativa para un grupo de escuelas con fines de supervisión e inspección administrativa. De manera similar, los municipios son una de las tres unidades jurisdiccionales básicas de gobierno y pueden contribuir en forma significativa a la infraestructura y las condiciones materiales de las escuelas.
6. Por ejemplo, *Framework for Teaching*, de C. Danielson, Perrenoud (2004); *Rewards and Incentives Group* (2009); Ministerio de Educación de Ontario (2009); Khim Ong (2008) y Ministerio de Educación de Singapur (2006).
7. Con respecto a las cantidades apropiadas para los incentivos, una revisión de los programas internacionales muestra que los incentivos para los maestros individuales pueden oscilar entre menos de 1% y más de 360% del sueldo mensual (OECD, 2009), aunque los expertos sugieren que entre 4% y 8% del sueldo anual puede ser adecuado para que los incentivos sean significativos pero no causen un comportamiento no deseado.

CAPÍTULO 1

México responde a los desafíos educativos

Con el propósito de abordar los persistentes retos en el rubro de la educación, el gobierno mexicano estableció prioridades claras en su *Programa Sectorial de Educación 2007-2012* (SEP, 2007). Para supervisar el avance de México hacia la consecución de los objetivos, el Programa estableció indicadores de mejora del logro del alumno tal como son medidos por la Evaluación Nacional del Logro Académico en Centros Escolares (ENLACE) y el Programa para la Evaluación Internacional de Alumnos de la OCDE (PISA, por sus siglas en inglés) (SEP, 2007).¹ Otros indicadores básicos se refieren al desarrollo profesional de los docentes, al reforzamiento de responsabilidades en la escuela, a la igualdad de oportunidades educativas y a las reformas relacionadas con el contenido y los planes de estudio. Como señal de su compromiso con los procesos de reforma, el gobierno mexicano estableció en 2008 la *Alianza por la Calidad de la Educación* con el sindicato nacional de maestros, lo que permitió definir los temas centrales del Acuerdo de Cooperación con la OCDE (SEP, 2008). La finalidad del Acuerdo fue determinar no sólo los cambios de políticas que México necesitaba, sino también la forma de diseñar e implementar reformas de políticas con eficacia, tomando en cuenta las condiciones, las limitantes y las oportunidades locales.

Un punto del Acuerdo con la OCDE fue la elaboración de políticas y prácticas adecuadas para evaluar la calidad de escuelas y docentes, así como para vincular los resultados del aprendizaje a incentivos con el fin de fomentar una mejora continua. Esta labor quedó a cargo del Consejo Asesor de la OCDE sobre Políticas de Evaluación e Incentivos Docentes en México, conformado por expertos invitados y personal asignado de la OCDE. El Consejo Asesor se concentró en primer lugar en los incentivos docentes, debido a que para la Secretaría de Educación Pública (SEP) esta era una de sus mayores prioridades.

Cuadro 1.1

Áreas cubiertas por la Alianza por la Calidad de la Educación y el Acuerdo OCDE-México

Alianza por la Calidad de la Educación	Acuerdo OCDE-México
Modernización de los centros	Gestión escolar y participación social
Profesionalización de docentes y autoridades educativas	Selección de los docentes
Bienestar y desarrollo integral de los alumnos	Trayectorias de carrera
Preparación de los alumnos para la vida y el trabajo	Incentivos y estímulos para los docentes
Evaluación para mejorar la calidad de la educación	Evaluación

Nota: El texto en azul indica los puntos del Acuerdo de Cooperación en los que se centró el Consejo Asesor de la OCDE sobre Políticas de Evaluación e Incentivos Docentes en México.

Fuente: SEP, 2008.

Cuadro 1.2

Metodología de la OCDE para apoyar la implementación de políticas educativas y resultados del Acuerdo de Cooperación

Metodología	Resultados
Análisis comparativo	<ul style="list-style-type: none"> – Visitas de campo por parte de personal de la OCDE a los estados mexicanos de Nuevo León, Chiapas, Aguascalientes y Veracruz. – Documentos de trabajo e informes elaborados por expertos invitados sobre: los incentivos docentes desde una perspectiva comparativa, la prueba ENLACE, marcos de políticas públicas y procesos de reformas, la rendición de cuentas, los resultados del aprendizaje del alumno para medir las mejoras, la evaluación de docentes en servicio, los incentivos docentes, las pruebas piloto, el monitoreo y la evaluación. – Revisión estatal sobre prácticas de incentivos docentes en México.²
Recomendaciones	– Recomendaciones específicas para México en las áreas mencionadas.
Comunicación	– Talleres internacionales, reuniones técnicas, presentaciones y actos de divulgación.

Con el fin de promover un diseño e implementación eficaces de políticas, el Secretariado de la OCDE asignó a varios miembros de su personal para participar en el Acuerdo de Cooperación, incluyendo a una persona en la ciudad de México para mejorar la comunicación e interacción entre los distintos actores. El Consejo Asesor fue conformado por expertos nacionales e internacionales en el área de la educación, quienes elaboraron una serie de recomendaciones sobre políticas educativas con los aportes del Secretariado. A lo largo del proceso hubo oportunidades para la consulta con actores clave del sistema educativo en México.

Este informe presenta las principales conclusiones y recomendaciones sobre políticas públicas propuestas por el Consejo Asesor de la OCDE sobre Políticas de Evaluación e Incentivos Docentes en México y el Secretariado de la OCDE como parte de los resultados finales del Acuerdo de Cooperación. El informe recoge los resultados de los talleres internacionales y las reuniones técnicas con los actores implicados en el sistema educativo mexicano, las visitas de campo, los reportes temáticos de expertos invitados y la investigación e información disponible en la OCDE. El informe y las publicaciones relacionadas forman parte de un conjunto mayor de obras que incluyen las distintas conclusiones y recomendaciones del Consejo Asesor de la OCDE sobre Gestión Escolar y Política Docente. Las otras publicaciones son:

- *Mejorar las escuelas: Estrategias para la acción en México*, derivado del trabajo del Consejo Asesor de la OCDE sobre Gestión Escolar y Política Docente;
- *Evaluación y reconocimiento de la calidad de los docentes: Prácticas internacionales*, editado por Susan Sclafani para el Consejo Asesor de la OCDE sobre Políticas de Evaluación e Incentivos Docentes en México; y
- *La medición del aprendizaje de los alumnos: Mejores prácticas para evaluar el valor agregado de las escuelas*, que actualizó y editó especialmente el Consejo Asesor de la OCDE sobre Políticas de Evaluación e Incentivos Docentes en México en el marco del Acuerdo de Cooperación.

Además de presentar las principales conclusiones y recomendaciones, este informe pretende facilitar que el gobierno mexicano mantenga los procesos de políticas educativas en curso dentro del marco de las reformas.³ La OCDE reunió abundante información durante la realización del Acuerdo y efectuó consultas frecuentes con los actores relevantes mexicanos. Para la OCDE, este Acuerdo significó un enriquecedor esfuerzo conjunto, la oportunidad de cooperar con un país miembro con el fin de apoyar sus esfuerzos para realizar reformas necesarias, y de implicarse directamente en un análisis en profundidad con los actores implicados.

El objetivo de las conclusiones y las recomendaciones de este informe es establecer un marco adecuado para la evaluación e incentivos docentes. En el capítulo 2 se ofrece un panorama del conocimiento actualmente disponible para los responsables de tomar decisiones respecto a las acciones para reformar la educación. El capítulo presenta un marco de políticas públicas para ayudar a los responsables de formular políticas en México y otros países a diseñar, planear, implementar y evaluar iniciativas en este rubro. El capítulo 3 explora la rendición de cuentas como eje de las políticas públicas para mejorar los resultados del aprendizaje estudiantil, y propone elementos específicos que deben considerar las autoridades educativas y los actores implicados. El capítulo 4 presenta un panorama de las medidas más comunes de los resultados del aprendizaje, al mismo tiempo que sugiere vincular un enfoque complementario que combine múltiples fuentes de evidencia con medidas cuantitativas. Para ilustrar este punto, el capítulo proporciona una visión general de los principales puntos fuertes, desafíos y oportunidades que representa la prueba ENLACE en México, y concluye con recomendaciones específicas para su fortalecimiento y desarrollo a futuro. El capítulo 5 contiene un panorama de los beneficios, desafíos y problemas de implementación que presentan los modelos de valor agregado enfocados en el aprendizaje y los logros del alumno, utilizados para mejorar las escuelas e incrementar la rendición de cuentas. El capítulo 6 bosqueja por un lado los problemas comunes y los desafíos de implementación respecto a la evaluación de docentes en servicio, y por el otro plantea recomendaciones específicas para México. A partir de ejemplos internacionales, el capítulo 7 presenta una amplia discusión de los principales retos de los incentivos docentes, proporcionando prácticas recomendadas para el monitoreo y la evaluación de este tipo de políticas. Cada capítulo comienza con un análisis general, pertinente para los responsables de formular políticas en contextos múltiples, y concluye con consideraciones y recomendaciones específicas para México.

ANTECEDENTES

El sistema educativo de México experimentó un acelerado crecimiento desde 1950 al hacer de la cobertura universal su prioridad durante la mayor parte del siglo XX. Las dimensiones de México y sus características geográficas convierten la tarea de educar a la población en un reto formidable. Una cuarta parte de la población total se encuentra en el grupo de edad normativo que debe asistir a la educación básica. El sistema de educación básica⁴ es amplio y se divide por tipo de escuela de acuerdo con el tamaño y ubicación de la comunidad, y en el caso de las comunidades indígenas por el origen étnico (INEE, 2010b) (cuadros 1.3 y 1.4).

Cuadro 1.3

Dimensiones del sistema de educación básica, año lectivo 2008-2009 (formal)

	Alumnos 25.6 millones (76% del total)	Docentes 1.1 millones (66% del total)	Escuelas 222 000 (89.5% del total)
Preescolar	18%	19%	40%
Primaria	58%	49%	44%
Secundaria	24%	32%	16%

Fuente: INEE, 2010b.

Cuadro 1.4

Educación básica por tipo de escuela, año lectivo 2008-2009

Preescolar	Primaria	Secundaria
General 89%	General 94%	General 51%
Indígena 8%	Indígena 6%	Técnica 28%
Comunitaria 3%	Comunitaria 1%	Telesecundaria 20%
		Comunitaria 0.3%

Fuente: INEE, 2010b.

La desigualdad es una realidad para muchos niños en México. Por ejemplo, en todo el país, el 92.4% de los niños ingresa a la escuela a la edad adecuada, aunque esto varía según el estado: en Chiapas la proporción cae a un 82.7%, comparada con el 96.6% en Aguascalientes.⁵ De igual modo, por cada 1 000 alumnos que en Chiapas ingresaron a la educación básica a la edad adecuada en el año lectivo 2000-2001, sólo 476 egresaron a tiempo seis años después (año lectivo 2007-2008), frente a los 747 niños de Aguascalientes (INEE, 2010b). Asimismo, hay claras diferencias entre los tipos de escuelas. De cada 1 000 alumnos, 198 abandonan la secundaria general al cabo de cinco años, mientras que la cifra es de 241 para las secundarias técnicas⁶ (INEE, 2010b).

El lugar en donde viven alumnos, su origen étnico y la pobreza son otros factores asociados a la desigualdad en la matrícula escolar en México. En escuelas primarias generales, los niños que ingresan a la escuela a una edad más tardía⁷ representan un 4.6% de la población escolar, cifra que se dispara a un 15.1% en las escuelas indígenas. Las diferencias entre escuelas en áreas urbanas y rurales, así como entre los distintos contextos socioeconómicos, son claras: 2.8% de los alumnos tiene mayor edad en las escuelas de áreas urbanas con poca pobreza, 6.9% en escuelas de áreas urbanas pobres, 5.6% en áreas rurales pobres y 11.5% en áreas rurales con pobreza extrema.⁸

Del mismo modo, existen retos en cuanto a la calidad de la educación entre y al interior de las mismas escuelas. Los resultados de PISA colocan a México entre los países de la OCDE con menor desempeño. Si bien se han registrado numerosas iniciativas de política pública en el sector educativo desde la década de los 90, los resultados no siempre han sido los deseados (OCDE, 2010b). Al principio de esa década, México inició un

rápido proceso de reformas tanto en la gobernanza como en el sector público. Temas como la transparencia, una sociedad informada y una mayor rendición de cuentas por parte del sector público comenzaron a integrar la agenda pública. Esto sugiere que México tiene una importante oportunidad para sentar las bases que auspicien acciones permanentes para reformar la educación. Una mayor rendición de cuentas por parte de escuelas, docentes y autoridades educativas, contar con un contenido y estándares de desempeño bien definidos, incrementar el liderazgo escolar, tener un marco educativo eficaz y políticas docentes amplias, son elementos que pueden contribuir a que los alumnos mejoren en su aprendizaje.

En México, la evaluación es una práctica relativamente reciente (Zorrilla, 2003). En 1993 se implementó algo parecido a una evaluación magisterial, el Programa Carrera Magisterial, cuya primera finalidad fue crear una fuente suplementaria de ingresos al salario de los docentes, y la segunda, ofrecerles la posibilidad de avanzar en la profesión. Es sólo a partir de 2006 que se aplican a los alumnos pruebas estandarizadas de tipo censales. El análisis reflejado en este informe contempla estos importantes avances y evalúa las oportunidades que tiene México para mejorar la calidad de la educación en el país, en particular en las áreas de evaluación de alumnos⁹ y del docente.¹⁰ La creación del Instituto Nacional de Evaluación Educativa (INEE) en 2002 respondió a la creciente demanda social de contar con un organismo independiente que efectuara evaluaciones confiables del sistema educativo (INEE, 2006).

Además de los beneficios que resultan de un mayor logro educativo tanto para los individuos como para las sociedades (OCDE, 2010a), una de las principales urgencias en México es garantizar el derecho a una educación de buena calidad para todos los alumnos. En este contexto, también son importantes las medidas y políticas de rendición de cuentas basadas en una mayor transparencia y evidencias. Sin embargo, aún hay mucho por hacer para que México garantice el derecho a una educación de buena calidad para todos los alumnos. Una gran cantidad de estos problemas por resolver tiene que ver con factores socioeconómicos (INEE, 2010a):

- Las poblaciones que reciben servicios educativos distintos presentan marcadas diferencias sociales, económicas y culturales.
- El acceso a la educación sigue siendo un problema pendiente del sistema de educación básica de México:
 - Una de las dificultades es que México no cuenta con mediciones directas de asistencia estudiantil.
 - En 2008, de cada 100 niños entre 6 y 11 años de edad, 98 asistieron a la escuela; de entre 12 y 14 años, sólo 92 de cada 100; y de entre 15 y 17 años, sólo 65 de cada 100.
 - Los niños de entre 4 y 6 años de edad, y de entre 15 a 17, corren un mayor riesgo de no asistir a la escuela si provienen de familias con menos recursos.
 - El trabajo infantil de más de 20 horas a la semana sigue siendo una causa de ausentismo escolar.
 - El ausentismo es más común en las áreas rurales que en las urbanas.
 - La pobreza persiste como obstáculo para alcanzar una cobertura universal. Mientras más marcada es la pobreza, mayor es la cantidad de menores que no asiste a la escuela.
 - Mientras más pequeña es la localidad, mayor es el porcentaje de niños que no va a la escuela.
 - El origen étnico también incide en la asistencia escolar.
 - En los niños de 12, 13, 14 y 15 años de edad, el avance disminuye y se considera insatisfactorio. Aunque no tan marcado, este fenómeno también se observa en la educación primaria.
- La calidad y pertinencia del aprendizaje es importante para cumplir con los requerimientos de un mundo en constante cambio:
 - Al final de la educación secundaria, 7 de cada 10 alumnos no cumplen con los objetivos educativos establecidos en el currículo nacional (medidos con EXCALE).
 - El logro de los objetivos educativos está relacionado estrechamente con el tipo de escuela a la que asiste el alumno.

- El material y los recursos humanos varían también según la ubicación y el tipo de escuela:
 - En la educación primaria, la proporción de escuelas comunitarias con al menos una computadora para uso educativo es de 0.5%, de 23.1% en escuelas indígenas y de 55.7% en escuelas generales.
 - La proporción de escuelas primarias urbanas con electricidad es del 99.5%, del 81.1% en escuelas indígenas y del 51% en comunitarias.
 - Las escuelas con los mayores porcentajes de docentes que cuentan sólo con estudios de educación media o incluso menos son las escuelas primarias comunitarias e indígenas.

El reto no es sencillo y será necesario encontrar soluciones en cada etapa de la cadena de la prestación del servicio educativo. Se requiere una visión de largo plazo que dirija un proceso gradual y que desde un principio empiece a dar pasos en la dirección correcta.

Al principio del Acuerdo de Cooperación, la SEP solicitó el apoyo de la OCDE para determinar la mejor manera de recompensar a los docentes conforme a su desempeño. Esto requirió una revisión exhaustiva de la experiencia y evidencia internacionales para identificar cuáles eran las mejores prácticas existentes que se podían aplicar en México. Los resultados se presentan en el libro disponible en español y en inglés *Evaluación y reconocimiento de la calidad de los docentes: Prácticas internacionales (Evaluating and Rewarding the Quality of Teachers: International Practices)* (OCDE, 2009), que forma parte del Acuerdo de Cooperación. Además, fue necesario analizar algunos elementos de la profesión docente como un todo, incluyendo la selección de candidatos desde la formación inicial hasta la jubilación, poniendo un énfasis especial en los docentes en servicio. A partir de la evidencia internacional queda claro que la evaluación sirve como instrumento para generar mejoras, y cada vez hay más conciencia entre los distintos actores en México de que el país necesita avanzar en esa dirección. Con el fin de profundizar más en lo anterior, se encargaron análisis a expertos centrados en la evaluación de docentes en servicio, modelos de evaluaciones sumativas y formativas, mecanismos de rendición de cuentas y estándares para la educación básica. Los resultados se presentan en este informe.

Con la creación del INEE, México ha podido recopilar de manera sistemática los resultados de pruebas, además de otros indicadores de la calidad de la educación como son la cobertura, la eficiencia, las tasas de abandono y de repetición de cursos. Este amplio volumen de datos y conocimiento sobre la evaluación es una base significativa para la rendición de cuentas. Sin embargo, México necesita mejorar todavía la calidad de los indicadores que emplea no sólo para la educación, sino también para los campos social y económico. La calidad de las pruebas estandarizadas requiere mejoras importantes para garantizar que los resultados puedan ser comparados y sustentar análisis más detallados. De igual modo, México necesita crear nuevos instrumentos para medir los resultados y el desarrollo educativo de alumnos, complementarios a las pruebas estandarizadas.

Una parte considerable de los análisis del Consejo Asesor y de los expertos invitados se centró en el sistema ENLACE y en las oportunidades y desafíos para su desarrollo. De esto se consideró que ENLACE es el mayor y más importante ejercicio evaluativo en el país dada la cantidad de grupos que cubre y los años durante los cuales se ha aplicado. Los conjuntos de distintos datos que se generan de esta forma tienen el potencial de fomentar nuevos avances en las políticas públicas. Los expertos invitados realizaron detallados y rigurosos análisis técnicos de las características de ENLACE. Los resultados muestran que ENLACE es un instrumento sólido para medir el logro del alumno. Una de las recomendaciones en torno a ENLACE es su fortalecimiento y mayor desarrollo. Mientras México mantiene estos esfuerzos para alcanzar las reformas educativas de mayor impacto, convendría hacer suyas esta y otras recomendaciones sugeridas en este informe.

NOTAS

1. Es importante observar que el primer objetivo mencionado en el documento de planeación del gobierno es mejorar la calificación media de México en PISA para pasar de 392 (promedio de las dos calificaciones medias para México en matemáticas y lectura) a 435 en 2012. El segundo indicador tiene que ver con el porcentaje de alumnos que alcanza un nivel satisfactorio en ENLACE (SEP, 2007, p. 15).
2. Se pueden encontrar los análisis y recomendaciones específicas sobre políticas docentes y liderazgo escolar en la publicación hermana de la OCDE *Mejorar las escuelas: Estrategias para la acción en México*. En esta obra se describen también algunas características del sistema educativo mexicano.
3. Las autoridades mexicanas emprendieron varias iniciativas como son el Programa de Transparencia y Rendición de Cuentas, el Programa de Estímulos a la Calidad Docente, las Escuelas de Calidad, así como el Sistema Nacional de Información de las Escuelas. México también trabaja actualmente con la OCDE para revisar metodologías de elaboración de indicadores y su comparación a nivel internacional.
4. La educación básica se divide en nivel preescolar (tres años), educación primaria (seis años) y educación secundaria (tres años).
5. Grupo de 1994 que finalizó en 2000.
6. Años lectivos de 2003-2004 a 2007-2008.
7. Dos años, o más, por encima de la edad normativa por grado y nivel educativo.
8. Año lectivo 2008-2009.
9. ENLACE y EXCALE son las principales pruebas estandarizadas que se aplican en México a los alumnos. Los objetivos, y lo que examinan estas pruebas, son distintos. ENLACE se aplica de manera amplia como prueba censal, y EXCALE se basa en muestras de alumnos.
10. La Carrera Magisterial es un sistema de promoción, basado en ciertas características de los docentes, que ha resultado ser ineficaz para mejorar los resultados de aprendizaje de alumnos. La participación de los docentes en la Carrera Magisterial es voluntaria.

Referencias bibliográficas

INEE (2006), *Policies and Systems for the Assessment of Education: Achievements and Challenges*, México.

INEE (2010a), *El derecho a la educación en México*, Informe 2009, México.

INEE (2010b), *Panorama educativo 2009*, México.

OCDE (2010a), *The High Cost of Low Educational Performance: The Long-Run Economic Impact of Improving PISA Outcomes*, OCDE, París.

OCDE (2010b), *Mejorar las escuelas: Estrategias para la acción en México*, OCDE, París.

SEP (2007), *Programa Sectorial de Educación—Secretaría de Educación Pública 2007-2012*, Gobierno de México, consultado el 24 de abril de 2010 en www.sep.gob.mx/sep1/programa_sectorial.

SEP (2008), *Alianza por la Calidad de la Educación*, consultado el 24 de abril de 2010 en <http://alianza.sep.gob.mx>.

Zorrilla, M. coordinadora (2003), *La evaluación de la educación básica en México: Una mirada a contraluz*, Universidad Autónoma de Aguascalientes, México.

CAPÍTULO 2

Marco de políticas públicas para implementar las reformas educativas

2.1 Reformas de políticas educativas en un contexto internacional...	28
2.2 Movilización de la investigación de la OCDE, de las prácticas internacionales y del conocimiento nacional.....	30
2.3 Ejes de política pública para las reformas educativas: La importancia de articular las preguntas correctas	33
2.4 Consideraciones para México	34

En esta era de economías basadas en la información y el conocimiento, los sistemas educativos de todo el mundo enfrentan el reto de proporcionar una educación de buena calidad a todos los ciudadanos como medio que abra oportunidades para las generaciones presentes y futuras por igual. Sin embargo, con la crisis económica global, los gobiernos deben afrontar ahora este desafío con políticas que permitan un financiamiento público sustentable sin dejar de promover el crecimiento y desarrollo económicos a largo plazo. Es por esto que los sistemas educativos se encuentran bajo una presión mayor para mejorar su desempeño en términos del aprendizaje de los alumnos, la equidad de oportunidades educativas y el valor de la inversión pública en la educación (OCDE, 2010a).

En este contexto, y con el fin de contribuir al debate en curso sobre las reformas de políticas educativas eficaces y eficientes, este capítulo propone un marco para que los países y jurisdicciones ajusten y diseñen las importantes etapas iniciales de sus respectivos procesos de reformas. También se describe un proceso basado en la experiencia, el análisis y la aplicación de la movilización del conocimiento para conjuntar las prácticas y experiencias internacionales con el diseño y la implementación eficaces de políticas públicas (OCDE, 2003, 2007a). El capítulo abre con la revisión de algunas de las principales tendencias de reformas educativas en un contexto internacional. Después identifica algunos de los retos comunes que enfrentan los sistemas educativos, y reconoce la aceptación creciente que el monitoreo y la evaluación de los resultados educativos deberán de incluir comparaciones internacionales. Al recurrir a la experiencia y las prácticas internacionales de las investigaciones de la OCDE y otras fuentes, el capítulo presenta prioridades comunes de políticas que han sido útiles para reformar políticas educativas, y que serán abordadas en los siguientes capítulos.

No obstante, los países pertenecientes a la OCDE como los asociados siguen investigando no sólo en *qué* puntos deben enfocarse los cambios de políticas sino también *cómo* se deben diseñar y aplicar con eficacia las reformas educativas dependiendo de las condiciones, limitaciones y oportunidades locales. El capítulo identifica un elemento que suele pasar inadvertido pero que es crucial para que los países adapten e implementen reformas de manera más eficiente y eficaz: un conocimiento confiable basado en experiencias específicas del país. Al combinarse con las prácticas internacionales y las experiencias comparables internacionalmente, la movilización del conocimiento local puede proporcionar un vínculo vital para adaptar las mejores prácticas y reformar eficazmente la educación, ajustándose a las prioridades y contextos nacionales. El *Marco de Políticas Públicas para Implementar Reformas Educativas* propuesto, plantea algunas áreas de políticas clave que deben ser consideradas en la movilización y análisis del conocimiento específico del país, después de que las políticas prioritarias hayan sido identificadas por los gobiernos y los actores implicados.

2.1 REFORMAS DE POLÍTICAS EDUCATIVAS EN UN CONTEXTO INTERNACIONAL

Para el bienestar actual y futuro de sus ciudadanos, las naciones deben atender las demandas de un contexto internacional competitivo y en constante cambio. Es por esto que tanto las economías desarrolladas como las que aún se encuentran en vías de desarrollo, requieren sistemas educativos que proporcionen a los alumnos las habilidades y conocimientos necesarios para ser eficaces e innovadores, no sólo en contextos locales o nacionales, sino también en economías globalizadas. Si bien la calidad y la equidad en la educación requieren esfuerzos e inversiones considerables, la mejora en los resultados educativos incrementa de forma drástica los beneficios para las finanzas públicas, las sociedades y los individuos (OCDE, 2010a, 2010b; Hanushek y Woessmann, 2009a).

Los modelos económicos recientes que establecen una relación entre las habilidades cognitivas y el crecimiento económico, muestran que pequeñas mejoras en la educación básica reportan grandes beneficios. Por ejemplo, con una modesta mejora de 25 puntos en PISA en los próximos 20 años, incluso un país con un desempeño educativo excelente como Finlandia se beneficiaría de un incremento del producto interno bruto (PIB), en términos del valor actual, de 553 mil millones de dólares estadounidenses, mientras que los países de bajo desempeño como México aumentarían su PIB en unos 5 billones de dólares estadounidenses (OCDE, 2010b, p.23). Asimismo, la equidad de oportunidades en la educación resulta vital para el futuro

crecimiento económico. Si los países garantizaran un desempeño de sus alumnos en los niveles mínimos de PISA (400 puntos) – inclusive aquellos con antecedentes socioeconómicos desfavorables y de diferentes grupos étnicos – los beneficios serían aun mayores: casi 15 billones de dólares estadounidenses para Turquía, 72 billones para los Estados Unidos de América y 26 billones para México (OCDE, 2010b, p. 26).

Los beneficios también son patentes en los niveles superiores de educación. Por ejemplo, el rendimiento público neto promedio de una educación superior (86 404 dólares estadounidenses para un varón) en los países de la OCDE casi triplica la inversión pública (OCDE, 2010a, indicador A8). Además de otros beneficios sociales referentes a la salud, el empleo, el compromiso público y la cohesión social, una mayor educación conlleva beneficios durante toda la vida de los individuos. Por ejemplo, contar con educación superior en los países de la OCDE representa un rendimiento financiero neto, durante la vida económicamente activa de un individuo, de más de 145 000 dólares estadounidenses para varones y de 92 000 para mujeres (OCDE, 2010a, indicador A8).

Por el contrario, también son elevados los costos de un desempeño educativo bajo y las correspondientes oportunidades perdidas para el bienestar social y el desarrollo económico. Así sucede en economías tanto en desarrollo como desarrolladas. Por ejemplo, en América Latina, una de las regiones de menor desempeño en evaluaciones educativas internacionales, la falta de calidad y equidad educativas explica una parte de su limitado desarrollo económico y social (Hanushek y Woessmann, 2009b). Para los países en el otro extremo del espectro del PIB, los cálculos muestran que si en 1998 el logro de los alumnos en los Estados Unidos de América hubiese sido el mismo que el de países como Finlandia y Corea del Sur, su PIB en 2008 habría sido 2.3 billones de dólares estadounidenses más elevado (McKinsey y Company, 2009).

Así, además de evaluaciones nacionales y locales, los países prestan cada vez más atención al desempeño y características de sus sistemas educativos en comparación al de otras naciones, lo cual es posible gracias a instrumentos internacionales como, entre otros, PISA (desde 2000) y la Encuesta Internacional sobre Docencia y Aprendizaje (*Teaching and Learning International Survey*, TALIS por sus siglas en inglés) (desde 2009).¹ Los estudios más recientes sobre reformas en la educación destacan la necesidad de que los países y sistemas educativos comparen su eficacia y eficiencia con estándares internacionales, basados en los de los países de mayor desempeño (OCDE, 2010c; ECS, 2008; McKinsey y Company, 2007). El uso de referencias y comparaciones internacionales (*international benchmarking* en inglés) en la educación implica que los aspectos específicos de cada nación, como estándares, prácticas docentes, desarrollo profesional, evaluación e incentivos, deben retomar las mejores prácticas internacionales (OCDE, 2009a; ECS, 2008).²

Las reformas en la educación ciertamente son vitales, pero al mismo tiempo representan un desafío. Reformas educativas anteriores y actuales, llevadas a cabo en varias economías desarrolladas y en desarrollo tanto dentro y fuera de la OCDE (*i.e.* Japón, México, los Estados Unidos de América, Brasil y China) se han enfrentado con resistencias y polémicas. Además de carecer de resultados, entre las deficiencias de las acciones de reformas comúnmente citadas se menciona que las reformas educativas reflejan prioridades a corto plazo en lugar de mejoras a largo plazo, que se implementan de manera bastante apresurada y sin la suficiente experimentación y adaptación locales, sin una evaluación eficaz ni sistemas de medición, y que excluyen a importantes actores como son los maestros y los padres de familia (*Canadian Council of Learning*, 2009; OCDE, 2009a). Además, algunos factores que contribuyen a la complejidad particular de las reformas educativas son los siguientes:³

- Inversión amplia: el gasto público promedio en instituciones educativas en los países de la OCDE fue de 6.2% del PIB, con un aumento de inversión pública y privada en todos los niveles educativos entre 1995 y 2007 de al menos 8% en términos reales (OCDE, 2010a, indicador B2).
- Tamaño del sector: el de los maestros es un grupo profesional particularmente extenso que representó en 2007 cerca del 4% del empleo en países de la OCDE (OCDE, 2010c). No sorprende que el mayor sindicato de América Latina, por ejemplo, sea uno magisterial (Santibáñez y Jarillo, 2007).

- Casi universal: la educación es una experiencia que casi todo el mundo tiene en grados diversos, y la enorme magnitud de servicios que ofrece cada día (OCDE, 2010a, indicador C1) lo hace uno de los sectores socialmente más visibles.

Estos factores, además de las condiciones, intereses y limitaciones locales propias de un país determinado, muestran que los intentos de introducir reformas en el sector educativo, incluso modestas, no son algo sencillo.

Sin embargo, la experiencia internacional también demuestra que es posible mejorar. Países como Finlandia, Japón, Corea del Sur y Polonia ilustran que es posible obtener resultados de aprendizaje elevados y equitativos, así como lograr mejoras rápidas (OCDE, 2001, 2004b, 2007d; McKinsey y Company, 2007). Los ejemplos de reformas educativas logradas también dejan ver que la tendencia internacional se aleja de la idea de aumentar únicamente la cobertura de servicios o de centrarse en los insumos educativos, para también garantizar la calidad (*i.e.* Brasil y México) (Hanushek, 2003), para enfocarse en los resultados de las inversiones en la educación (*i.e.* los Estados Unidos de América). El interés internacional en los resultados del aprendizaje del alumno como criterio clave para medir el éxito, destaca de igual forma otro punto: el creciente gasto público no se traduce necesariamente de manera directa en mejores resultados del aprendizaje. Por ejemplo, si bien los Estados Unidos de América triplicaron el gasto promedio real por alumno desde 1960 (Hanushek, 2003), el logro educativo está estancado tanto en PISA como en las evaluaciones nacionales (OCDE, 2001, 2004b; NCES, 2003).

Las reformas educativas dependen hoy día de presupuestos públicos limitados, con cargas de déficit incluso en economías ricas que oscilan en torno a un 7.8% del PIB en 2010 (OCDE, 2010d). En plena crisis económica global, la eficiencia de los sistemas educativos, así como su eficacia, seguirán siendo una prioridad. Para países con un déficit creciente y presupuestos públicos limitados, sencillamente no será posible aumentar en un futuro cercano los salarios de todos los maestros para que, por ejemplo, correspondan al estatus y prestigio que gozan en países con sistemas educativos de mejor desempeño (*i.e.* Corea del Sur, Irlanda y Singapur), que también presumen contar con los profesionales docentes mejor pagados (OCDE, 2010a, indicador D3; Sclafani, 2008; Murphy y Coolahan, 2003). Cada vez más se presta una mayor atención para implementar por un lado formas distintas de incentivar a los maestros para que mejoren sus prácticas de enseñanza y consigan que todos sus alumnos obtengan niveles más elevados en el aprendizaje; por otro lado, se intenta atraer y retener a los mejores y más brillantes candidatos para la profesión docente (OCDE, 2009a).

Sin embargo, si los países y sistemas educativos han de adaptar e implementar políticas de reformas basadas en las mejores prácticas internacionales, las condiciones, limitantes y oportunidades locales deben abordarse de manera adecuada. Los estudios de procesos de reformas en múltiples regiones, países y contextos llegan a conclusiones semejantes (OCDE, 2010c, 2009a, 2007a; The World Bank, 2008a, 2008b; IDB, 2006, 2008, 2010). No obstante, el proceso para llevar a cabo las reformas es todavía un asunto mayoritariamente de ensayo y error para gobiernos y responsables de tomar decisiones.

2.2 MOVILIZACIÓN DE LA INVESTIGACIÓN DE LA OCDE, DE LAS PRÁCTICAS INTERNACIONALES Y DEL CONOCIMIENTO NACIONAL

Tenemos una ventaja comparativa para distinguir el “qué” de las reformas.

[...] Pero ahora se nos pide que también ayudemos a los países con el “cómo” de las reformas...

Angel Gurría,
Secretario general de la OCDE
Septiembre de 2008, México⁴

Los países que llevan a cabo reformas educativas centradas en el desempeño del alumno, en la equidad de oportunidades educativas⁵ y en un mayor valor de la inversión, no sólo deben determinar las prioridades de políticas que necesitan adoptar, sino también cómo implementarlas. Los responsables de formular políticas públicas y los actores implicados revisan las prácticas y experiencias de investigaciones internacionales en busca

de ayuda para diseñar, planear e implementar reformas de políticas públicas. Las organizaciones internacionales, como es el caso de la OCDE, reciben cada vez más solicitudes de países miembros y socios para que se les proporcionen los análisis de políticas educativas y los procesos de reformas más recientes y relevantes.

Conforme la OCDE y otros órganos internacionales de expertos intentan responder con eficacia a tales demandas de orientación, deben filtrar y sintetizar una cantidad significativa de información, datos y conocimientos recientes. Con investigaciones originales y el trabajo de proyectos específicos, los conocimientos disponibles sobre educación se pueden clasificar como sigue: perfiles nacionales con base en indicadores y comparaciones internacionales (*i.e.* ediciones anuales de *Education at a Glance*); evaluaciones y estudios internacionales (*i.e.* PISA, TALIS, el Programa para la Evaluación Internacional de Competencias de Adultos, PIAAC por sus siglas en inglés); investigación de temas específicos (*e.g.* sobre políticas docentes y formación vocacional, y sobre los costos de un bajo desempeño educativo), revisiones y notas temáticas de países específicos (*e.g.* educación de inmigrantes en Dinamarca); y revisiones de prácticas internacionales (*e.g.* *Evaluación y reconocimiento de la calidad de los docentes: Prácticas internacionales* y las mejores prácticas de modelos de valor agregado en las escuelas).⁶

A pesar de la considerable acumulación de conocimientos, los países miembros de la OCDE expresan la necesidad de tener acceso a una herramienta dinámica, simultánea y específica por nación, que conduzca a los responsables de formular políticas y a los profesionales en servicio a las estrategias de reformas, experiencias e implementaciones más actualizadas. Esta necesidad surge por la dificultad de establecer vínculos claros entre políticas y reformas con los resultados, en particular debido al lapso de tiempo relativamente extenso requerido para evaluar el impacto de los cambios en el sector educativo (OCDE, 2010c). Existen pocas conclusiones definitivas que resultan de la investigación sobre temas de políticas específicas, lo que aumenta la presión a la OCDE y otros organismos internacionales para aprovechar la cantidad acumulada de conocimientos y poner en primer plano las áreas donde se necesita realizar mayor trabajo de investigación (OCDE, 2007a). Las acciones más recientes en esta área al interior de la OCDE, reflejan no sólo las demandas de economías globalizadas que cambian con rapidez (*i. e.* *Estrategia de habilidades de la OCDE*), sino también las demandas de conocimientos de países que emprenden reformas (por ejemplo, *el Acuerdo de Cooperación para mejorar la calidad de la educación en las escuelas mexicanas* y *El aprovechamiento del conocimiento para mejores políticas educativas-Proyecto GPS*).

Las prácticas, investigaciones y experiencias internacionales, así como una serie de crisis económicas y políticas (OCDE, 2010c), se combinaron y conformaron una agenda de reformas educativas para la educación básica. Aunque haya todavía numerosos debates sobre aspectos específicos y se necesite investigar más, existe una convergencia sobre algunos elementos básicos relacionados con la calidad, la equidad y el valor de la inversión en educación. Según el contexto nacional y las prioridades locales, los elementos básicos de reformas emergentes se pueden agrupar en torno a:

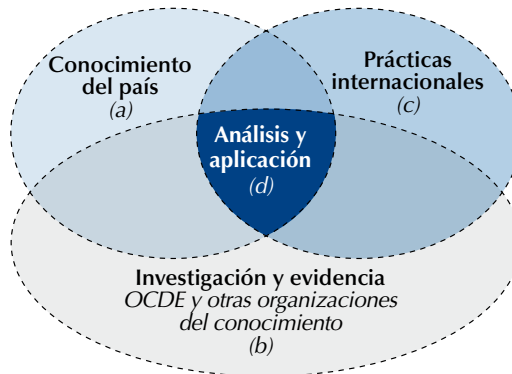
- La rendición de cuentas – centrándose en distintos niveles de la cadena de la prestación del servicio educativo (sistema, instituciones, escuelas y maestros, y alumnos).
- Los estándares, la evaluación y el currículum – incluyendo las medidas de aprendizaje y avance del alumno, los estándares de contenido y desempeño para alumnos y maestros, las evaluaciones formativas y sumativas, la alineación y la coherencia.
- Las políticas docentes – incluyendo la evaluación, el desarrollo profesional, los incentivos, la formación y la selección.
- El liderazgo escolar, la evaluación y la mejora de las escuelas – incluyendo la evaluación de contribuciones netas al aprendizaje del alumno (valor agregado), los temas de la autonomía, la administración, la participación de los padres de familia, la competencia y la elección de la escuela.
- Los incentivos y estímulos – incluyendo aquellos para las jurisdicciones (*i.e.* órganos estatales y financiamiento federal), así como para las escuelas y los maestros.

No obstante, una lista de prioridades y mecanismos – el *qué* de las posibles intervenciones – no basta para diseñar, planear e implementar reformas eficaces de política pública en una sola o en una combinación de estas áreas. Incluso cuando se cuenta con las mejores prácticas, investigaciones y experiencias de la OCDE y otras fuentes, las condiciones, capacidad y costos específicos de cada país pueden limitar en gran medida su utilidad (OCDE, 2010c, 2009a, 2007a; IDB, 2010, 2008, 2006; Spiller *et al.*, 2003; The World Bank, 2008a, 2008b). Los factores específicos de cada país revisten particular importancia para toda reforma educativa porque no hay un modelo universal, o un conjunto específico de instrumentos de política pública, que genere un resultado global consistente, ni siquiera en temas particulares como el de los instrumentos fiscales (OCDE, 2010c), de los mecanismos de rendición de cuentas,⁷ o de los incentivos y estímulos para los maestros (OCDE, 2009a). Además, el diseño, la consulta, la planeación, los programas piloto, la implementación y la evaluación rigurosa de las políticas, con frecuencia no siguen el mismo ritmo de los mandatos administrativos y los presupuestos limitados de la mayoría de los gobiernos. Los países de la OCDE y los socios reconocen de manera explícita la importancia de la investigación y los conocimientos locales para diseñar políticas basadas en la evidencia (OCDE, 2007a).⁸

Es por eso que el *cómo* de las reformas requiere un sólido modelo heurístico de procesos de reformas de política pública que tome en cuenta los conocimientos específicos de cada país, analizados junto con la investigación, las evidencias y las prácticas internacionales. Una representación esquemática del concepto de movilización, análisis y aplicación del conocimiento (*Knowledge MAP* por sus siglas en inglés) se presenta en la gráfica 2.1. De esta manera, los conocimientos combinados de diversas fuentes (*a*, *b* y *c* de la gráfica 2.1) sobre temas específicos de política pública se analizan, procesan y al final se aplican para conseguir una implementación eficaz y resultados sostenibles (*d* en la gráfica 2.1).

Gráfica 2.1

Representación esquemática del proceso de movilización, análisis y aplicación eficaces del conocimiento
Heurística específica para un país



En este modelo heurístico de movilización del conocimiento, el reto es ordenar los conocimientos pertinentes y específicos de la manera más adecuada y rentable con el fin de adaptar, diseñar e implementar políticas educativas apropiadas para el contexto nacional. Los análisis rigurosos de estos datos contribuyen a identificar las intervenciones y los métodos de implementación más adecuados que pueden adaptarse e incorporarse a las reformas de política pública en el ámbito nacional. Esto se da en el área *d* de la gráfica 2.1, donde los conocimientos basados en la experiencia de diversas fuentes se analizan y aplican a la luz de las realidades locales y las limitantes de tiempo de los mandatos administrativos, por ejemplo.

2.3 EJES DE POLÍTICA PÚBLICA PARA LAS REFORMAS EDUCATIVAS: LA IMPORTANCIA DE ARTICULAR LAS PREGUNTAS CORRECTAS

A pesar del amplio periodo de tiempo que se necesita para introducir las reformas de política pública, las administraciones gubernamentales trabajan en general con plazos de tiempo más cortos, por lo que a menudo se toman decisiones de política de forma un tanto apresurada, con poca evidencia y recursos, y en contextos difíciles – si no abiertamente adversos (OCDE, 2010c). Además, salvo que se planeen con antelación, las finanzas públicas rara vez permiten invertir grandes cantidades en estudios *ex ante* sobre temas específicos con el propósito de permitir decisiones mejor informadas sobre el diseño de políticas públicas. Aunque la mayoría de las naciones tiene en las distintas ramas del gobierno sistemas bastante sólidos y eficaces de control de las finanzas públicas (*i.e.* cuentas nacionales, registros legislativos, y entidades de estadísticas y de evaluación), la gran cantidad de registros, bancos de datos y estadísticas disponibles dentro del sector educativo y en todas las áreas relacionadas (*i.e.* en los ministerios, tanto horizontal como verticalmente), también puede dificultar la identificación de los datos y la información más pertinentes y útiles para tomar decisiones oportunas en el diseño de políticas públicas.

Por lo anterior, la OCDE ha elaborado un *Marco de Políticas Públicas para Implementar Reformas Educativas*, cuyo objeto no sólo es ayudar a que los países miembros ubiquen con exactitud los datos nacionales pertinentes sobre los temas vitales para el diseño de políticas públicas, sino que también identifiquen las áreas que quizá necesiten más estudio y sistematización. El Marco aborda seis ejes de política pública que facilitan identificar las condiciones, limitantes, oportunidades y costos en un país o jurisdicción particular. El Marco se debe aplicar a la prioridad específica, o un tema particular de las políticas públicas, que requiere decisiones instruidas sobre su diseño, planeación e implementación. Es necesario elaborar indicadores básicos en cada uno de los ejes propuestos para garantizar que la información recopilada sea pertinente y aplicable al tema específico de política pública que se analiza. En la medida de lo posible, estos indicadores deben asociarse y compararse con datos disponibles a nivel internacional de otros sistemas de alto desempeño para tener puntos de referencia y realizar comparaciones internacionales. El Marco ofrece a los gobiernos y a los actores locales un método para movilizar el conocimiento y los análisis locales en problemas particulares. Los seis ejes de las políticas reflejan los componentes clave que han demostrado facilitar u obstaculizar las reformas educativas (OCDE, 2010c, 2009a, 2007a; IDB, 2010, 2008, 2006; Spiller, Stein y Tommasi, 2003; The World Bank, 2008a, 2008b):

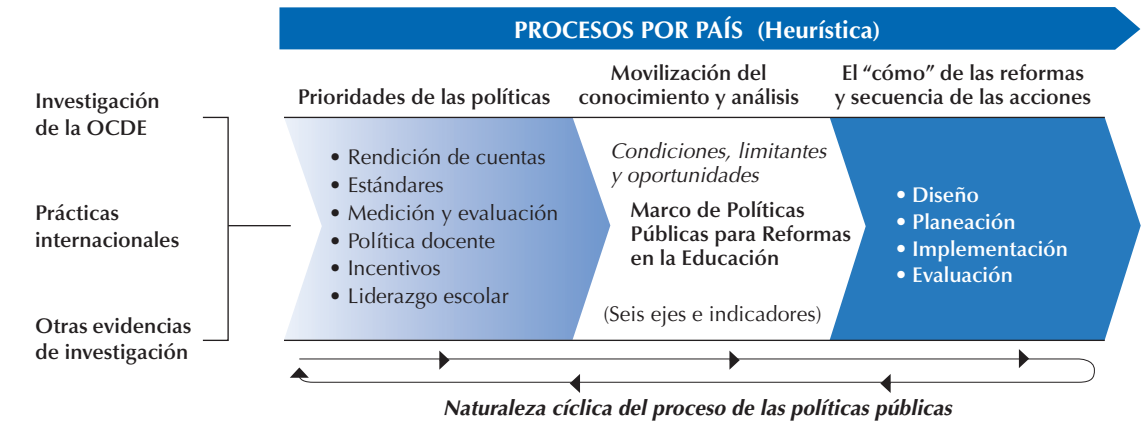
- i) La calidad y la cantidad de *la información y los datos pertinentes disponibles* (sobre alumnos y maestros, el desempeño, y posibles vínculos) para fijar líneas base, objetivos y metas. Es importante que se logren establecer líneas base referidas a los temas particulares sobre los cuales se pretende incidir con las reformas educativas.
- ii) *La estrategia de comunicación, compromiso y consulta* con los actores implicados (incluyendo al público en general, los líderes de opinión y las organizaciones profesionales y de maestros). Por ejemplo, las acciones de las reformas que no sean percibidas como pertinentes por los padres de familia, o por la sociedad en su conjunto, son más susceptibles de fracasar durante el proceso de implementación y conforme vayan surgiendo dificultades (OCDE, 2009a, 2008).
- iii) La cantidad y disponibilidad del *financiamiento público* para la elaboración e implementación de la reforma educativa (si se debe renovar anualmente o forma parte del presupuesto fijo), incluyendo un análisis de la inversión requerida, proyecciones de costos y posibles economías o reasignación de fondos de programas vigentes.
- iv) *Un marco legal, normativo y administrativo adecuado*, para identificar conflictos potenciales y posibles modificaciones que sean requeridas en el ámbito educativo y temas relacionados (por ejemplo, en las leyes laborales, la participación social y las leyes de información y privacidad).
- v) *Arreglo institucional* entre instituciones de carácter público con competencias relevantes (para la evaluación, el currículum, la investigación, etc.).
- vi) *Proceso de descentralización formal y de facto* entre las entidades responsables de los servicios educativos (incluyendo consideraciones de los recursos, servicios y capacidades institucionales y operativas).

Si bien los países pueden tener capacidades diferentes respecto a la generación, movilización y administración del conocimiento, a menudo es la información básica y disponible (“lo que está al alcance de la mano”), que resulta la más útil y menos costosa de recopilar. Los organismos académicos, gubernamentales o de cooperación internacional, pueden proporcionar ayuda a cualquier país o jurisdicción que la requiera para movilizar el conocimiento necesario, en particular en torno a métodos cuantitativos (OCDE, 2003). Con conocimientos actualizados, oportunos y confiables sobre temas específicos de política pública, cualquier país o jurisdicción puede llevar a cabo un sólido proceso de adaptación, diseño y planeación de políticas públicas. En combinación con otras fuentes de información y evidencias sobre políticas (prácticas internacionales, trabajos de investigación de la OCDE y otras fuentes), un diseño de políticas que tome en cuenta las mejores prácticas puede adaptarse a las condiciones, las limitantes, los costos y las oportunidades locales (gráfica 2.2).

Gráfica 2.2

**Representación esquemática de un modelo heurístico específico al contexto local de país:
Marco de Políticas Públicas para Implementar Reformas Educativas**

Desempeño, equidad y eficacia de la inversión en educación



El propósito del *Marco de Políticas Públicas para Implementar Reformas Educativas* es proporcionar a los gobiernos y a los distintos actores un método específico para movilizar el conocimiento en el ámbito local, centrado en las prioridades específicas identificadas. Lo que se pretende garantizar de este modo es que, tanto los responsables de formular políticas, como los actores implicados basen sus decisiones no sólo en las prácticas, la experiencia y el asesoramiento internacionales, sino también en el conocimiento local más actualizado y pertinente.⁹ Los seis ejes del proceso de las políticas públicas sirven también como criterios de evaluación para asegurar que las propuestas de políticas y programas que pretenden dar “respuestas” a los tomadores de decisiones se basen en el adecuado conjunto de preguntas (elaboradas a partir de cada uno de los seis ejes de política pública presentados). Al igual que en la investigación científica, la pertinencia de los resultados en la elaboración de políticas públicas está relacionada con iniciar el proceso a partir de las preguntas correctas.

2.4 CONSIDERACIONES PARA MÉXICO

En su *Programa Sectorial de Educación 2007-2012* (SEP, 2007), el gobierno mexicano definió prioridades claras de política pública para reformar la educación. Para supervisar el avance hasta alcanzar los objetivos previstos, el Programa estableció indicadores de mejora del rendimiento de alumnos, medidos conforme a

las evaluaciones de ENLACE y PISA. Es importante observar que la Secretaría de Educación Pública (SEP) ha emprendido varias iniciativas de reformas actualmente en curso, como se mencionó en el capítulo anterior. En este contexto, este capítulo y las consideraciones y recomendaciones que a continuación se presentan tienen como objetivo proporcionar a la SEP, y a los actores implicados en México, elementos para definir los pasos específicos a seguir con base en las prioridades educativas definidas, el conocimiento local, los factores limitantes y las oportunidades que existen en México para un proceso de reforma duradero.

- Además de las prácticas internacionales y los resultados de las investigaciones disponibles, se recomienda la **movilización del conocimiento específico del país** sobre temas de política pública como un elemento vital para diseñar, planear e implementar eficazmente reformas educativas viables y sustentables tomando en cuenta las condiciones, limitaciones y oportunidades que hay en México. La labor de la OCDE en esta área, así como la experiencia de diversas regiones en contextos múltiples, sugieren con claridad que un conocimiento confiable y actualizado sobre temas particulares de políticas públicas, es crucial en el proceso de adopción de las mejores prácticas y recomendaciones sobre políticas educativas en un contexto específico. La movilización del conocimiento enfocada en los ejes relevantes de políticas públicas puede darse de manera bastante rápida (de seis a 12 meses, por ejemplo) en el ámbito nacional. Este proceso de movilización del conocimiento en México puede contribuir en gran medida al análisis, al diseño y a la planeación de iniciativas de políticas educativas basadas en las mejores prácticas internacionales y evidencia disponible.
- Aunque hay varios aspectos de políticas públicas concernientes a las reformas que pretenden mejorar el aprendizaje del alumno y el desempeño escolar (e.g. estándares, evaluación, medición, incentivos, liderazgo escolar y el desarrollo profesional), **existen ejes comunes de políticas públicas que deberán guiar el proceso de movilización del conocimiento local**. Tales ámbitos ofrecen un marco con el cual la SEP y los actores del sector educativo en México pueden continuar el proceso de análisis, el diseño, la planeación, la implementación y la evaluación de políticas educativas. Las acciones actuales y futuras de reformas educativas en México podrán beneficiarse del conocimiento local actualizado en cada uno de los siguientes seis ejes de políticas:
 - i) La calidad y cantidad de datos e información pertinentes disponibles (sobre alumnos y maestros, desempeño y vínculos entre ellos) para fijar objetivos y metas e identificar áreas deficientes que se deben examinar.
 - ii) Las opciones de estrategias de comunicación, compromiso y consulta con los principales actores implicados, incluyendo al público en general, maestros, directores y autoridades educativas locales. Esto implica identificar cómo traducir las reformas propuestas en un mensaje socialmente pertinente y significativo para las familias, los maestros y los directores de escuelas.
 - iii) La cantidad y disponibilidad del financiamiento público para la elaboración e implementación de las reformas educativas, incluyendo el análisis de rentabilidad potencial, proyecciones de costos y economías que se obtengan al reorientar rubros o programas del presupuesto vigente.
 - iv) El arreglo institucional existente entre los organismos de carácter público, como por ejemplo entre la SEP, el INEE y las autoridades educativas estatales, para identificar organismos específicos que deban contribuir a la elaboración de estándares, evaluaciones y proponer modificaciones.
 - v) Un marco legal, normativo y administrativo actualizado, que resuelva los conflictos potenciales que surjan a partir de las modificaciones que se requieran en el ámbito educativo, pero también en rubros relacionados como la ley laboral y lo referente a la participación social y la gestión de recursos escolares.
 - vi) El proceso de descentralización *formal* y *de facto* de los principales organismos federales y estatales responsables de proveer los servicios educativos (esto incluye considerar los recursos, el manejo de bases de datos y de la información, la evaluación, la supervisión y la prestación de servicios educativos).

- Para informar y aconsejar a las autoridades educativas y a los actores implicados, la constitución de una **entidad incluyente y objetiva encargada de la gestión y el análisis del conocimiento** puede facilitar el proceso de elaboración de las políticas educativas. Las áreas responsables de analizar y diseñar las políticas públicas al interior de la SEP y en los organismos relacionados deben ser fortalecidas para contribuir a que esta entidad sea vista como un mediador a nivel nacional. No es necesario crear una nueva entidad oficial, lo que se requiere es designar un órgano que enfrente los retos que se suelen manifestar durante las reformas educativas:
 - La necesidad de contar con una revisión, análisis y opiniones sustentadas y oportunas, esbozadas por los propios expertos en el área, por las prácticas y las experiencias de investigaciones internacionales y nacionales disponibles.
 - La necesidad de contar con la opinión fiable y objetiva de expertos en torno al diseño y los resultados de los procesos de monitoreo y evaluación de programas de incentivos docentes y escolares, por ejemplo, así como de otros programas principales de las reformas educativas.
 - Los desfases temporales que existen entre las administraciones gubernamentales (federales, estatales y municipales) y los períodos fijos de mandato, y los requerimientos a largo plazo que exigen la planeación, los programas piloto, la evaluación y la investigación para sustentar reformas educativas. Las entidades pueden adoptar distintas formas (por ejemplo, consorcio, consejo, comité, o programa transversal), y las que se han establecido en Brasil, Canadá, Dinamarca, Singapur y los Estados Unidos de América pueden ofrecer ejemplos útiles para México.

NOTAS

1. Otros incluyen las Tendencias en el Estudio Internacional de Matemáticas y Ciencias (TIMSS, por sus siglas en inglés) y las acciones regionales como las evaluaciones UNESCO/LLECE (1997 y 2006) en 16 países de América Latina y el Caribe.
2. Como se analiza en los siguientes capítulos, los estándares proporcionan un método eficaz para garantizar que el contenido, el desempeño y las prácticas docentes en un sistema educativo sean internacionalmente competitivos.
3. Adaptado de “Making Reform Happen in Education” en OCDE, 2010c, p. 161.
4. Tomado del discurso de apertura del Secretario general de la OCDE, Angel Gurría, pronunciado el 25 de septiembre de 2008 en la Ciudad de México: *The Art of Making Reform Happen: Learning from Each Other*.
5. La importancia de la equidad en las acciones de la reforma sobresale en las investigaciones que muestran que, por ejemplo, a pesar de que las políticas de rendición de cuentas generan un aumento en los resultados del alumno, incluso en contextos de mucha polémica como el de los Estados Unidos de América, la mejora en el aprendizaje no es igual entre todos los grupos de alumnos (Hanushek y Raymond, 2006).
6. Otro ejemplo importante de este tipo de enfoque comparativo es el informe de 2007 de McKinsey & Company, en el que se destaca la importancia de fijar criterios internacionales para el desempeño del alumno.
7. La experiencia revela que los enfoques y el diseño de políticas públicas son importantes para alcanzar políticas eficaces de rendición de cuentas y para mejorar el aprendizaje del alumno (Hanushek y Raymond, 2006; Hanushek, Raymond y Rivkin, 2004). Son justamente las limitantes, el contexto y las oportunidades locales de un país o sistema educativo específicos, lo que determina en gran medida la especificidad del diseño de políticas públicas.
8. Es importante distinguir entre investigaciones llevadas a cabo con el propósito de informar sobre las acciones emprendidas (sobre las políticas basadas en la información) y la investigación puramente científica que, junto con diversos estándares y la demostración de efectos causales (OCDE, 2007a), pueden requerir períodos más largos para que los resultados sean visibles.
9. Esto también podría identificar las contradicciones posibles entre las prácticas internacionales y la información y la evidencia recabada localmente en el país. En estos casos, ahondar en el análisis sería un paso necesario y valioso para las autoridades educativas y los actores implicados.

Referencias bibliográficas

- Canadian Council on Learning** (2009), "Changing Our Schools: Successful Educational Reform", Ottawa, Ontario.
- Education Commission of the States (ECS)** (2008), *From Competing to Leading: An International Benchmark Blueprint*, Denver, Colorado.
- Hanushek, Eric A.** (2002), "The Long Run Importance of School Quality", documento de trabajo NBER núm. W9071, julio.
- Hanushek, Eric A.** (2003), "The Failure of Input-Based Schooling Policies", en *Economic Journal*, vol. 113, núm. 485, febrero, pp. F64-F98.
- Hanushek, Eric A., Margaret E. Raymond y Steven G. Rivkin** (2004), *Does It Matter How We Judge School Quality?*, documento presentado en los American Education Finance Association Annual Meetings, Salt Lake City, Utah, del 11 al 13 de marzo.
- Hanushek, Eric A. y L. Woessmann** (2009a), *Do Better Schools Lead to More Growth? Cognitive Skills, Economic Outcomes, and Causation*, National Bureau of Economic Research, documento de trabajo 14633, Cambridge, Massachusetts.
- Hanushek, Eric A. y L. Woessmann** (2009b), *Schooling, Cognitive Skills, and the Latin American Growth Puzzle*, National Bureau of Economic Research, documento de trabajo 15066, Cambridge, Massachusetts.
- Hanushek, Eric A. y Margaret E. Raymond** (2006), "School Accountability and Student Performance", en *Regional Economic Development*, vol. 2, núm. 1, pp. 51-61.
- Inter-American Development Bank (IDB)** (2006), *The Politics of Policies: Economic and Social Progress in Latin America*, Harvard University Press, Cambridge, Massachusetts.
- IDB** (2008), *Political Institutions, State Capabilities, and Public Policy: International Evidence*, Inter-American Development Bank, documento de trabajo, serie IDB-WP-661.
- IDB** (2010), *Political Institutions, Policymaking, and Economic Policy in Latin America*, Inter-American Development Bank, documento de trabajo, serie IDB-WP-158.
- King, E.M. y S. Cordeiro Guerra** (2005), "Education Reform in East Asia: Policy, Process, and Impact", en The World Bank, *East Asia Decentralizes: Making Local Government Work*, pp. 179-207.
- McKinsey & Company** (2007), *How the World's Best-Performing School Systems Come Out on Top*, Identifying Teacher Quality Project, Washington, DC.
- McKinsey & Company** (2009), *The Economic Impact of the Achievement Gap in American Schools*, McKinsey Social Sector Office, Nueva York.
- Murphy, I. y J. Coolahan** (2003), "Attracting, Developing and Retaining Effective Teachers: Country Background Report for Ireland", con autorización de la página principal de la OCDE, OCDE, París.
- National Centre for Education Statistics (NCES)** (2003), "NAEP Trends in Academic Progress Through 1999", en *NCES Digest of Education Statistics 2003*, NCES, Washington, DC.
- OCDE** (2001), *PISA 2000: Knowledge and Skills for Life*, OCDE, París.
- OCDE** (2003), *New Challenges for Educational Research*, OCDE, París.
- OCDE** (2004a), *Innovation in the Knowledge Economy: Implications for Education and Learning*, OCDE, París.
- OCDE** (2004b), *PISA 2003: Learning for Tomorrow's World*, OCDE, París.
- OCDE** (2007a), *Evidence in Education: Linking Research and Policy*, OCDE, París.

- OCDE** (2007b), *No More Failures: Ten Steps to Equity in Education*, OCDE, París.
- OCDE** (2007c), *Getting it Right: OECD Perspectives on Policy Challenges in Mexico*, OCDE, París.
- OCDE** (2007d), *PISA 2006: Science Competencies for Tomorrow's World*, OCDE, París.
- OCDE** (2008), *Measuring Improvements in Learning Outcomes: Best Practices to Assess the Value-Added of Schools*, OCDE, París.
- OCDE** (2009a), *Evaluating and Rewarding the Quality of Teachers: International Practices*, OCDE, París.
- OCDE** (2009b), *Education Today: The OECD Perspective*, OCDE, París.
- OCDE** (2010a), *Education at a Glance 2010: OECD Indicators*, OCDE, París.
- OCDE** (2010b), *The High Cost of Low Educational Performance: The Long-Run Economic Impact of Improving PISA Outcomes*, OCDE, París.
- OCDE** (2010c), *Making Reform Happen*, OCDE, París.
- OCDE** (2010d), *Economic Outlook No. 87*, OCDE, París.
- Santibáñez, L. y B. Jarillo** (2007), "Conflict and Power: The Teachers' Union and Education Quality in Mexico", en *Well-Being and Social Policy*, vol. 3, núm. 2, pp. 21-40.
- Sclafani, S.** (2008), "Rethinking Human Capital in Education: Singapore as a Model for Teacher Development", documento preparado para The Aspen Institute Education and Society Program, Aspen Institute, Washington, DC.
- Spiller, Pablo T., E. Stein y M. Tommasi** (2003), "Political Institutions, Policymaking Processes, and Policy Outcomes: An Intertemporal Transactions Framework", Latin American Research Network (IDB).
- The World Bank** (2008a), *Raising Student Learning in Latin America: The Challenge for the 21st Century*, Washington, DC.
- The World Bank** (2008b), *The Road Not Travelled: Education Reform in the Middle East and North Africa*, Washington, DC.
- Woessmann, L. y T. Fuchs** (2004), *What Accounts for International Differences in Student Performance? A Re-Examination Using PISA Data*, documento de análisis IZA núm. 1287, documento de trabajo CESifo, serie núm. 1235.

CAPÍTULO 3

La rendición de cuentas como impulsor de políticas para mejorar los resultados del aprendizaje

3.1 Tipos y características de los sistemas de rendición de cuentas en la educación.....	40
3.2 Desempeño docente: Hacia una comprensión más amplia de la rendición de cuentas.....	42
3.3 Consideraciones para México.....	44

Los gobiernos ponen cada vez más un mayor énfasis en los sistemas de rendición de cuentas como parte de reformas más amplias en la administración del sector público (OCDE, 2009a). En el sector educativo, esto se refleja en un desplazamiento del interés en los *insumos* y *procesos* para concentrarse en los *resultados* (Hanushek y Raymond, 2006; Hanushek, 2002). Como el propósito final de los sistemas educativos es el aprendizaje del alumno, esto implica que los actores deben rendir cuentas de los logros de los alumnos. De manera paradójica, una mayor rendición de cuentas también implica transferir una mayor autonomía a los actores para decidir cómo movilizar los recursos que se les entregan para alcanzar los objetivos sobre los cuales deberán rendir cuentas. En este sentido, la rendición de cuentas implica por una parte que los actores necesitan demostrar públicamente que en realidad persiguen las metas establecidas, y por la otra que reciben la capacitación, el apoyo y la autonomía suficientes para alcanzarlas.

El propósito de este capítulo es ofrecer, primero, un panorama de las principales características de los sistemas de rendición de cuentas en la educación, y mostrar de qué manera son útiles para apoyar acciones que mejoren el logro de los alumnos. La primera parte empieza por presentar los principales tipos y características de los sistemas de rendición de cuentas. Tomando en cuenta las características comunes de los sistemas eficaces de rendición de cuentas en la educación, es palpable la importancia que cobran los estándares, tanto de los contenidos como del desempeño.

A partir de ejemplos internacionales ilustrativos, como los de Australia, Canadá, Finlandia y Hong Kong-China, este capítulo defiende la necesidad de tener un concepto más amplio de rendición de cuentas que sirva de soporte para las acciones de mejoras. Los sistemas educativos eficaces van más allá de la noción estrecha de rendición de cuentas “basada en pruebas”, misma que ha sido asociada a una ola de reforma educativa basada en estándares (Darling-Hammond, 2004). Una noción más amplia de rendición de cuentas tiene aplicaciones directas en relación con el desempeño y el grado de esfuerzo realizado por los docentes. El capítulo concluye con consideraciones para las autoridades educativas y los actores implicados en las reformas en México, preocupados por aumentar el desempeño de los alumnos, las escuelas y los maestros mexicanos.

3.1 TIPOS Y CARACTERÍSTICAS DE LOS SISTEMAS DE RENDICIÓN DE CUENTAS EN LA EDUCACIÓN

La naturaleza y las características de la rendición de cuentas en la educación dependen de las metas establecidas por el sistema educativo. Así, hay al menos cinco amplias categorías de rendición de cuentas que pueden ser tomadas en cuenta (Darling-Hammond, 2004; Ladd, 2007; NAE, 2009):

- **Rendición de cuentas política:** los funcionarios electos y las autoridades educativas nombradas por estos funcionarios están sujetos a rendir cuentas ante los ciudadanos de las circunscripciones electorales que los llevaron al puesto (la reciente polémica sobre evaluación y los resultados educativos en la Ciudad de Nueva York a finales de 2010 es un ejemplo de ello).
- **Rendición de cuentas legal:** tanto escuelas, maestros y órganos locales, como consejos escolares, están obligados a ceñirse al marco legal y legislativo establecido.
- **Rendición de cuentas administrativa:** distintas jurisdicciones responsables de controlar, supervisar o proporcionar servicios educativos establecen procedimientos regulatorios y administrativos para garantizar que los servicios y las escuelas cumplan con el desempeño esperado (*i. e.* las acciones recientes de reformas en los Estados Unidos de América).
- **Rendición de cuentas profesional:** los maestros, directores y demás miembros del personal escolar deben cumplir con los requerimientos básicos referidos al conocimiento especializado, al ingreso a la profesión y a los estándares de la práctica docente y al desempeño esperado. En esquemas con este tipo de rendición de cuentas, los docentes se sienten responsables de forma “horizontal” ante sus compañeros docentes y directores de escuela, así como ante los padres de familia.

- **Rendición de cuentas ligadas a la oferta y demanda** (“market accountability”): esto se relaciona a menudo con el grado de elección escolar, el financiamiento y otras consecuencias para la escuela relacionadas con las preferencias de las familias y de los alumnos.

Dependiendo de los objetivos establecidos y las prioridades de las políticas, los sistemas educativos se enfocan en ciertos tipos de rendición de cuentas. Las prácticas de varios países sugieren que en años recientes las acciones de reformas comenzaron a diversificarse para incluir formas de rendición de cuentas no sólo legales y burocráticas sino también profesionales y, cada vez más, con enfoques de oferta y demanda (OCDE, 2005, 2007a; Baker, 2010).

Sin embargo, la evidencia y las experiencias internacionales sugieren que los sistemas de rendición de cuentas enfocados en el aprendizaje del alumno deben (OCDE, 2009a, 2007a; NAE, 2009; ECS, 2008):

- basarse en expectativas claras sobre los resultados deseados;
- emplear medidas adecuadas, tanto formativas como sumativas, para evaluar el avance hacia los objetivos;
- contar con sistemas de datos e información adecuados;
- incluir mecanismos de información y comunicación para la rendición de cuentas pública y la transparencia; y
- vincularse a acciones y consecuencias que resultan del desempeño.

Lo anterior implica que las expectativas claras de los sistemas educativos deben reflejarse en **estándares claros de contenido y desempeño para alumnos y maestros** (OCDE, 2009a; NAE, 2009). Además, las cuestiones de alineación curricular y coherencia entre los estándares, las evaluaciones y el desarrollo profesional de los maestros son importantes para garantizar que trabajen juntos todos los elementos implicados en la rendición de cuentas y los procesos de mejora (Baker, 2010; NAE, 2009). Esto implica que es importante manejar un enfoque complementario para la evaluación que combine medidas sumativas y formativas, como demuestran algunos de los sistemas de mejor desempeño, que es el caso de Australia, Finlandia y Alberta, Canadá (Darling-Hammond y McCloskey, 2008; *Government of Alberta (Canada) – Education*, 2009). Incluso cuando los estándares de contenido se elaboran después de que tengan lugar las evaluaciones, como sucedió en Australia con el Programa de Evaluación Nacional de Lectoescritura y Razonamiento Matemático anual (*National Assessment Program in Literacy and Numeracy*) y el Currículo Nacional de Australia (*Australia National Curriculum*), se deben alinear el currículum (contenido), los estándares de desempeño, las prácticas de enseñanza y las evaluaciones (VCAA, 2010). En un interesante ejemplo de cooperación entre sistemas educativos en relación a las mejores prácticas en evaluación, Hong Kong-China y Queensland (Australia) tienen enfoques semejantes y han compartido lecciones aprendidas (Darling-Hammond y McCloskey, 2008). En Finlandia, el currículum nacional proporciona directrices para que los maestros elaboren evaluaciones al interior de las escuelas sin tener que recurrir a evaluaciones externas para clasificar o comparar a los alumnos o a las escuelas. En este contexto, el cuadro 3.1 presenta un resumen de los tipos de evaluación más frecuentes y sus usos en estos y otros sistemas de rendición de cuentas.

No obstante, dentro de los sistemas de rendición de cuentas, las evaluaciones externas son importantes. Los análisis en múltiples niveles de los resultados de PISA 2006 muestran una correlación positiva entre el logro del alumno y la publicación de los resultados de la escuela, así como con la existencia de exámenes externos basados en estándares (OCDE, 2007b). En cuanto a los sistemas adecuados de datos e información, los mecanismos de reporte y las acciones relacionadas con el desempeño, los capítulos 4, 5, 6 y 7 de esta publicación ofrecen un panorama de prácticas recomendadas y ejemplos internacionales. Sin embargo, pese a los beneficios de una mayor rendición pública de cuentas centrada en el desempeño del alumno, las nociones estrechas que reducen la rendición de cuentas a políticas unidimensionales “basadas en pruebas” han limitado su potencial (Ladd, 2007). Por tal motivo, la siguiente sección sugiere la necesidad de contar con una noción más amplia de rendición de cuentas.

Cuadro 3.1

Tipos de evaluación más comunes usados en sistemas de rendición de cuentas

Tipo de evaluación	Propósitos	Medidas	Fortalezas	Debilidades o inconsistencias potenciales
Evaluación formativa al interior del salón de clases	Mejorar el aprendizaje del alumno con base en información de diagnóstico obtenida según necesidades al interior del salón de clases.	Elaborada por el maestro, currículum integrado con diagnóstico detallado. Calificadas por el maestro.	Pertinente para la enseñanza; usa los conocimientos del maestro; potencialmente de apoyo a la práctica docente; permite utilizar respuestas elaboradas por alumnos con mayor validez.	Depende de la capacidad del maestro para diseñar evaluaciones, interpretar los resultados y aplicar las correcciones adecuadas. Se adapta a los alumnos pero tal vez no sea útil para comparar los resultados; puede crear dificultades para manejar el salón de clases.
Evaluación formativa de procedencia externa	Mejorar el aprendizaje de los alumnos y brindar información de diagnóstico.	De procedencia externa; calificada por el maestro o de manera externa.	Libera al maestro de los requerimientos de diseño de evaluaciones; está bien alineada con los estándares; puede permitir respuestas abiertas.	Puede ser inoportuna para muchos alumnos; puede no coincidir con los contenidos. Puede ser que los maestros no la comprendan bien.
Evaluación sumativa	Para informar sobre la situación actual y el avance de los alumnos.	De procedencia externa, por lo general calificada por maestros distintos a los de los alumnos.	Bajo costo si respuestas son cerradas. Confiable; puede servir en modelos de crecimiento y de valor agregado.	Transferencia deficiente; resultados poco confiables por la enseñanza de la prueba. Alineación deficiente; poca validez.
Evaluación combinada del salón de clases y sumativa para la rendición de cuentas	Para informar sobre la situación actual.	Elementos ponderados de evaluaciones formativas, y un componente de procedencia externa.	Incluye evaluaciones más válidas; alineada; permite transferencia.	Alto costo de implementación. Conflicto de intereses entre los maestros si se utiliza para medir el desempeño docente. Porcentaje problemático de ponderación.

Fuente: Baker, 2010.

3.2 DESEMPEÑO DOCENTE: HACIA UNA COMPRESIÓN MÁS AMPLIA DE LA RENDICIÓN DE CUENTAS

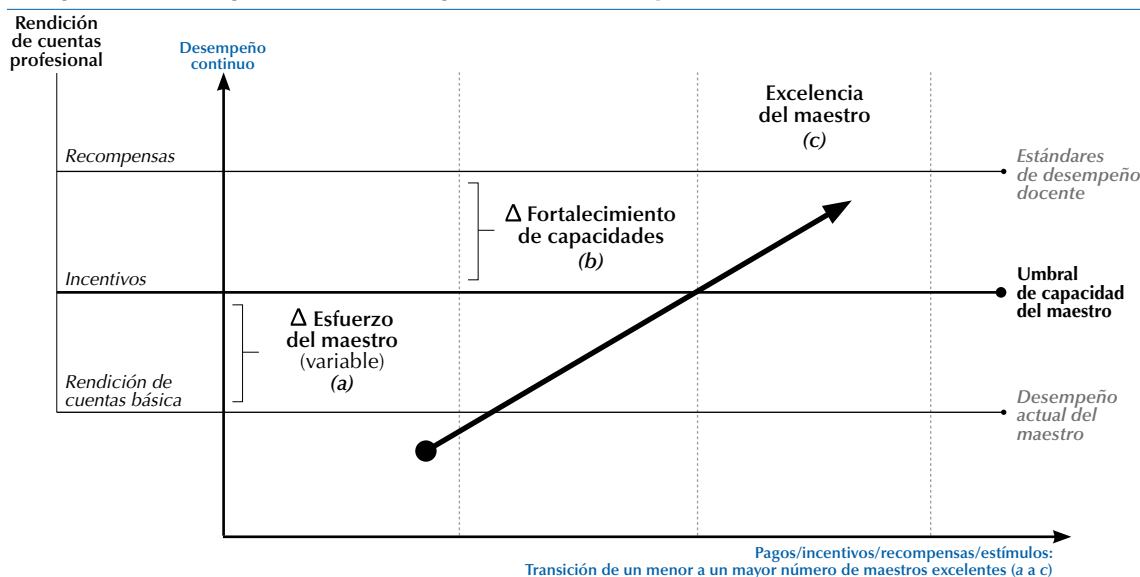
Con base en el análisis previo se puede entender que la rendición de cuentas comprende tres aspectos de un mismo concepto. En primer lugar, la rendición de cuentas en un sistema educativo dado puede referirse a una definición compartida y a una responsabilidad colectiva por el aprendizaje y el avance de los alumnos. Este aspecto de la rendición de cuentas se relaciona con la importancia y el valor que se brindan de manera clara a la educación en un país, y con el grado de prioridad del que goce en la sociedad y en la agenda pública (Baker, 2010). En segundo lugar, la rendición de cuentas también se refiere a los sistemas “con consecuencias”, en los que pueden haber consecuencias significativas para los individuos, las escuelas o jurisdicciones al formar parte del sistema de rendición de cuentas (OCDE, 2009a). Por ejemplo, las reformas en los Estados Unidos de América ya son sinónimo del debate entre jurisdicciones y actores sobre la rendición de cuentas (Ladd, 2007). Sin embargo, existe un tercer aspecto de la rendición de cuentas que se puede considerar “de consecuencias mínimas” o rendición de cuentas *básica*. Este tipo de rendición de cuentas se asocia con el grado de esfuerzo mostrado por los maestros y personal de la escuela para cumplir con los estándares mínimos de comportamiento profesional, tales como llegar a la escuela, llegar a tiempo y concentrarse en las actividades adecuadas. Dada la importancia de los maestros para el logro del alumno (OCDE, 2009a), es útil que los sistemas no sólo pongan atención en la rendición de cuentas basada en la escuela como unidad, sino también en la rendición de cuentas de docentes conforme a estándares profesionales.

Dentro de un sistema educativo que ha establecido estándares de desempeño para los maestros, el enfoque común es ofrecer oportunidades a todos los maestros para desarrollar las capacidades que les permitan alcanzar el desempeño previsto. Se pueden proporcionar diferentes formas de incentivos a los maestros para

motivarlos a que busquen oportunidades de generar aptitudes y tengan mejores posibilidades de alcanzar los estándares (Lavy, 2009). A los maestros que rebasen los estándares de desempeño se les puede ofrecer una mayor compensación para que incluso los maestros con alto desempeño tengan incentivos para que sigan mejorando su práctica (OCDE, 2009a).

En un sistema convencional en donde hay escasa rendición de cuentas, la teoría de contrato sienta las bases del problema típico de la asimetría de información entre los maestros y las familias o los responsables de formular las políticas públicas, generando un riesgo moral (Ferreya y Liang, 2010). Una de las suposiciones más comunes e implícitas de los sistemas de evaluación docente enfocados en la evaluación formativa, la creación de capacidades y el desarrollo profesional, es que el *desempeño* docente real es igual o muy cercano a la *capacidad* del docente. Sin embargo, la realidad a la que se enfrenta la mayoría de los sistemas educativos es que algunos maestros no se esfuerzan lo suficiente, desempeñándose muy por debajo de sus verdaderas capacidades. Aunque estos maestros de bajo desempeño (por falta de esfuerzo) pueden aprovechar y beneficiarse de las oportunidades de capacitación y desarrollo profesional, el escenario más eficiente es aquel en que se reduce la diferencia entre el esfuerzo y la capacidad del maestro (“ Δ Esfuerzo del maestro” en la gráfica 3.1).

Gráfica 3.1

Representación esquemática del concepto del *umbral de capacidad* en la rendición de cuentas docente

Fuentes: Elaborado a partir de OCDE (2009a) y Murphy (2001).

Los maestros que se desempeñan por debajo de su capacidad natural deberían recibir, en teoría, sólo el salario básico obligatorio (área “a”). Para los maestros que se desempeñan al límite de su capacidad natural, los incentivos y opciones de capacitación deberán permitirles, en teoría, aumentar su capacidad y alcanzar los estándares de desempeño establecidos por el sistema y superarlos, en su caso. Sin embargo, surgen ineficiencias cuando los maestros que deberían alcanzar o superar los estándares de desempeño de manera natural, mediante su esfuerzo y capacidades personales, incluso sin tener que aumentar sus capacidades, no lo consiguen. En este caso, la rendición de cuentas tendría que ser lo bastante fuerte para hacer que el maestro se motive a trabajar con el esfuerzo adecuado (es decir, esforzarse más) y que también busque diversas oportunidades para aumentar su capacidad.

El escenario óptimo en términos de recursos y aprendizaje de alumnos es contar con la mayor cantidad posible de maestros que se desempeñen en el umbral de su capacidad. En teoría, esto por sí solo aumentaría el desempeño docente promedio y el desempeño de alumnos de manera automática, sin que el sistema necesite aumentar las capacidades o proporcionarles otros estímulos para este fin. Con esta base de desempeño (es decir, el concepto del umbral de su capacidad), los maestros seguirán siendo motivados para mejorar su desempeño mediante la capacitación y el desarrollo profesional. En este modelo simplificado, la distancia entre el desempeño promedio de los maestros y los estándares de desempeño variará dependiendo de la calidad general de la fuerza docente. La utilidad de este modelo es significativa al menos por tres razones:

i) Como se analiza en el capítulo 6, una diferencia entre otros sistemas y los sistemas educativos con poca rendición de cuentas (*i. e.* maestros con plaza permanente a quienes no puede separárseles de la profesión independientemente de su desempeño y los resultados de aprendizaje de los alumnos), como los hay en América Latina, es que además de atraer y conservar en la profesión a los más brillantes y mejores maestros (como lo hacen Singapur y Suecia, por ejemplo), **es muy importante mejorar y apoyar a los maestros en servicio.**

ii) Con un mayor compromiso público y sistémico con el aprendizaje y mejoras de los alumnos, el desempeño docente se vuelve más importante. En un contexto de mayor rendición de cuentas, se espera que los maestros realicen un esfuerzo mayor para cumplir con una conducta profesional mínima y esperada de todos los docentes.

iii) **Las conductas de los maestros que se relacionan más con el esfuerzo que con la capacidad, como son la asistencia, la puntualidad y el tiempo efectivo dedicado a la enseñanza, tienen efectos importantes en el aprendizaje de los alumnos.** Algunos estudios en los Estados Unidos de América y en países en vías de desarrollo muestran que las ausencias de los maestros tienen efectos sorprendentemente altos en el aprendizaje de los alumnos, en particular en los alumnos de las comunidades más pobres y remotas. Por ejemplo, Das *et al.* (2007) reportan que en escuelas primarias en Zambia un aumento de 5% en las ausencias del maestro redujo el aprendizaje anual de un alumno promedio entre 4% y 8%. La evidencia sugiere de igual modo que es importante identificar el margen de error y las discrepancias en los registros oficiales. La experiencia en Ecuador e India, por ejemplo, indica que los datos oficiales difieren de las observaciones de campo sobre la asistencia/ausencia del maestro entre un 7% y un 25%. Esto es particularmente significativo para ciertos estados y regiones de México.¹ Los resultados de la Encuesta Internacional sobre Docencia y Aprendizaje (TALIS por sus siglas en inglés) de la OCDE revelan que en México, el ausentismo y el tiempo de trabajo efectivo de los maestros son retos importantes (OCDE, 2009b).

Así, el modelo basado en el concepto del *umbral de capacidad docente* permite asociar claramente la rendición de cuentas con los incentivos docentes en un desempeño continuo. Además, al aumentar la rendición de cuentas en un sistema para que los maestros trabajen a su capacidad máxima, puede haber posibles ganancias sustanciales en el desempeño promedio del maestro y el aprendizaje de los alumnos. Esto tiene implicaciones directas en las políticas públicas tanto en el diseño de los criterios de evaluación docente, como en el de las medidas del desempeño docente y escolar. Por ende, la meta final de un sistema de rendición de cuentas que incluya incentivos basados en el desempeño sería, por ejemplo, hacer que tantos maestros como sea posible pasen primero del punto “a” al “b” y después al “c” de la Gráfica 3.1. Este proceso se vería reflejado en el desempeño de los sistemas educativos conforme logren pasar de “deficiente” a “bueno” y de “bueno” a “excelente” en términos de los logros de sus alumnos.

3.3 CONSIDERACIONES PARA MÉXICO

Con base en las consideraciones anteriores y en las condiciones, limitantes y oportunidades específicas de México respecto a la rendición pública de cuentas enfocada en el logro de alumnos (Salieri *et al.*, 2010), las siguientes consideraciones se dirigen a las autoridades educativas y a los actores implicados en las reformas educativas.

- **La importancia de la rendición pública de cuentas: todos los actores implicados deben sentirse responsables y estar sujetos a rendir cuentas públicamente en torno al aprendizaje de los alumnos y de los resultados educativos globales.** Los actores deben estar sujetos a rendir cuentas del aprendizaje y desarrollo de los alumnos, considerando la asistencia y la capacitación necesarias. Un sistema de rendición de cuentas bien definido y enfocado en los resultados del aprendizaje y el incremento de los logros de los alumnos, puede proporcionar la coherencia general que se necesita dado el tamaño, la complejidad y los distintos intereses de los múltiples actores del sector educativo en México.

Usar el aprendizaje del alumno como criterio clave a partir del cual las autoridades educativas estatales, escuelas, directores y maestros deberán rendir cuentas, refleja la tendencia de enfocarse en los resultados del sistema educativo en lugar de los insumos brindados. Por ejemplo, la práctica internacional relacionada con los incentivos basados en el desempeño docente refleja este cambio. Sin embargo, esto no implica restar importancia a los problemas de infraestructura o inclusión social en el sistema educativo mexicano. En cambio, sí implica que el aprendizaje y la formación de todos los alumnos – es decir, que sean fomentados, cultivados, valorados y evaluados por distintos mecanismos – sean las metas finales de las acciones y reformas de políticas educativas. El apoyo a alumnos, escuelas, directores y maestros, así como el desarrollo profesional, son elementos vitales para una mayor rendición de cuentas.²

- **La rendición de cuentas enfocada en el aprendizaje y el desarrollo del alumno implica establecer estándares claros.** La elaboración de estándares como componente clave del sistema de rendición de cuentas enfocado en el aprendizaje del alumno debe abordar al menos tres prioridades: *i*) elaboración adecuada de estándares para el contenido y el desempeño del alumno y del maestro; *ii*) alineación y coherencia entre estándares, evaluación, medición y desarrollo profesional; y *iii*) alineación de estándares con las mejores prácticas internacionales y los criterios competitivos internacionalmente sobre los conocimientos y habilidades de los alumnos. Dentro de un marco de rendición de cuentas basado en estándares, los actores deben tener incentivos para alcanzar, o inclusive hasta rebasar, las expectativas reflejadas en los estándares.
- **Las medidas de rendición de cuentas deben incluir criterios complementarios de esfuerzo y desempeño.** Un sistema de rendición de cuentas basado en estándares para los alumnos y los maestros en México debe considerar el uso de medidas de aprendizaje y desarrollo del alumno (a partir de evaluaciones estandarizadas y otros métodos confiables, en la medida de lo posible), así como criterios complementarios referentes al desempeño individual, grupal y escolar. Esto es importante en México porque la asistencia, la puntualidad y el tiempo de trabajo efectivo de alumnos y maestros son aún temas importantes por resolver. Un sistema de rendición de cuentas en México debe asumir que algunos directores y maestros no se desempeñan a toda su capacidad. Por lo tanto, se necesitan incentivos para aumentar el esfuerzo y el desempeño básicos, y para incrementar las capacidades y el desarrollo profesional. Adicionalmente, la reducción del porcentaje de alumnos que dejan la escuela puede considerarse como un indicador adicional. La rendición de cuentas también implica que se puede aconsejar a los maestros que reciben asistencia técnica adecuada y oportunidades de desarrollo profesional, pero que son incapaces de mejorar su desempeño, que se retiren de la profesión docente.
- **La atención debería enfocarse en los alumnos, escuelas y maestros para conseguir una mejora continua, tomando la escuela como una unidad básica para la rendición de cuentas.**³ Si bien distintos niveles y actores del sistema educativo deben estar sujetos a la rendición de cuentas, la escuela es útil como unidad básica para la rendición de cuentas, pues cuenta con datos individuales, información y supervisión de los alumnos y los maestros. Los datos e información de alumnos y maestros a nivel escolar sirven para apoyar las acciones de mejora, los incentivos y estímulos docentes, intervenciones educativas para quienes presenten un bajo desempeño, y para identificar buenas prácticas y posibles modelos para elaborar estándares docentes, por ejemplo.

- **Es importante definir un proceso gradual para elaborar enfoques complementarios de evaluación a partir de múltiples fuentes de evidencia.** Elaborar un sistema sólido de rendición de cuentas basado en estándares es un proceso gradual, con etapas claras y enfoques complementarios de evaluación y medición. Las evaluaciones tanto sumativas como formativas del aprendizaje y crecimiento del alumno, así como del desempeño escolar y docente, deben formar parte del sistema de rendición de cuentas en México.

NOTAS

1. Alcázar *et al.* (2006) reportan que en Perú los índices de ausentismo de los maestros aumentan en un promedio de 11% a 16% y 21% al analizar datos de escuelas alejadas y de bajos recursos. Los datos de escuelas primarias en Bangladesh, Ecuador, India, Indonesia, Papúa Nueva Guinea, Perú, Uganda y Zambia muestran que las ausencias de los maestros van de un 11% en Perú a un 27% en Uganda (citado en Rogers y Vegas, 2009).
2. Por ejemplo, se pueden encontrar recomendaciones referentes al apoyo, creación de capacidades y desarrollo profesional para maestros en la publicación hermana de la OCDE *Mejorar las escuelas: Estrategias para la acción en México*.
3. La unidad de rendición de cuentas se refiere al nivel con base en el que se supervisan, miden y evalúan el esfuerzo, las capacidades y el desempeño de estudiantes, maestros y directores. Si bien los estudiantes son evaluados de manera individual, y los maestros deberían estar motivados individualmente en mejorar sus prácticas, al final los resultados deben agruparse de tal modo que los individuos se sientan responsables de manera colectiva a nivel de cada escuela.

Referencias bibliográficas

Alcázar, L., H. Rogers, N. Chaudhury, J. Hammer, M. Kremer y K. Muralidharan (2006), "Why are Teachers Absent? Probing Service Delivery in Peruvian Primary Schools?", en *International Journal of Educational Research*, vol. 45, pp. 117-136.

Baker, E.L. (2010), "Summative and Formative Evaluation in Educational Accountability", documento de trabajo preparado para las labores del Consejo Asesor sobre Evaluación y Políticas de Incentivos, UCLA, California.

Chaudhury, N., J. Hammer, M. Kremer, K. Muralidharan y H. Rogers (2006), "Missing in Action: Teacher and Health Worker Absence in Developing Countries", en *Journal of Economic Perspectives*, vol. 20, núm. 1, pp. 91-116, consultado el 12 de abril de 2009 en www.typon-link.com/AEAP/doi/pdf/10.1257/089533006776526058?cookieSet=1.

Darling-Hammond, L. (2004), "Standards, Accountability, and School Reform", en *Educational Leadership*, 55(5), 6-11.

Darling-Hammond, L. y L. McCloskey (2008), "What Would it Mean to Be 'Internationally Competitive'?", Stanford University, consultado el 10 de mayo de 2010 en http://edpolicy.stanford.edu/pages/events/kerner/materials/intnl_assessment_pdk.pdf.

Das, J., S. Dercon, J. Habyarimana y P. Krishnan (2007), "Teacher Shocks and Student Learning – Evidence from Zambia", en *The Journal of Human Resources*, vol. 42, núm. 4, pp. 820-862.

Education Commission of the States (ECS) (2008), *From Competing to Leading: An International Benchmark Blueprint*, Denver, Colorado.

Ferreira, M.M. y P.J. Liang (2010), "Information Asymmetry and Equilibrium Monitoring in Education", Carnegie Mellon University.

Glewwe, P., N. Ilias y M. Kremer (2008), "Teacher Incentives in Developing Countries: Recent Experimental Evidence from Kenya", documento de trabajo 2008-09, National Center on Performance Incentives, consultado el 15 de enero de 2010 en www.performanceincentives.org/data/files/directory/ConferencePapersNews/Glewwe_et_al_2008.pdf.

Government of Alberta (Canada) – Education (2009), "The Alberta Student Assessment Study: Final Report", Edmonton, Alberta.

Hanushek, Eric A. (2002), "The Long Run Importance of School Quality", documento de trabajo NBER núm. W9071, julio.

Hanushek, Eric A. y Margaret E. Raymond (2006), "School Accountability and Student Performance", en *Regional Economic Development*, vol. 2, núm. 1, pp. 51-61.

Hanushek, E.A. y M.E. Raymond (2006), "School Accountability and Student Performance", en *Regional Economic Development*, Federal Reserve Bank of St. Louis, vol. 2, núm. 1, pp. 51-61.

Hong Kong Examinations and Assessment Authority (2010), *Press Release: 2010 Hong Kong Advanced Level Examination Results Released*, consultado el 20 de junio de 2010 en www.hkeaa.edu.hk/DocLibrary/Media/PR/20100629_HKALE_Results_ENG.pdf.

Ladd, H.F. (2007), "Holding Schools Accountable Revisited", Association for Public Policy Analysis and Management, Washington, DC.

Murphy, K. (2001), "Performance Standards in Incentive Contracts", en *Journal of Accounting and Economics*, vol. 30, pp. 245-278.

Muralidharan, K. y V. Sundararaman (2009), "Teacher Performance Pay: Experimental Evidence from India", documento de trabajo 15323, National Bureau of Economic Research, consultado el 15 de diciembre de 2009 en www.nber.org/papers/w15323.pdf.

National Academy of Education (NAE) (2009), "Standards, Assessments, and Accountability – Education Policy White Paper", Washington, DC.

OCDE (2005), *Teachers Matter: Attracting, Developing and Retaining Effective Teachers*, OCDE, París.

OCDE (2007a), *Education at a Glance 2007: OECD Indicators*, OCDE, París.

OCDE (2007b), *PISA 2006: Science Competencies for Tomorrow's World*, OCDE, París.

OCDE (2009a), *Evaluating and Rewarding the Quality of Teachers* (S. Sclafani, ed.), OCDE, París.

OCDE (2009b), *Creating Effective Teaching and Learning Environments: First Results from TALIS*, OCDE, París.

Parandekar, S., E. Amorim y A. Welsh (marzo de 2008), "Prova Brasil: Building a Framework to Promote Measurable Progress in Learning Outcomes", en *En breve*, 121, pp. 1-4.

Rogers, F.H. y E. Vegas (2009), *No More Cutting Class? Reducing Teacher Absence and Providing Incentives for Performance*, The World Bank Policy Research, documento de trabajo 4847, The World Bank, Washington, DC.

Salieri, G., L. Santibáñez y B. Naranjo (2010), *State-Level Teacher Evaluation and Incentive Practices in Mexico: Diagnostic Study*, estudio encargado por la OCDE para el Acuerdo de Cooperación entre la OCDE y el gobierno de México.

The State of Queensland: Department of Education and the Arts (2005), *Queensland Curriculum, Assessment and Reporting Framework*, Brisbane, Strategic Policy and Education Futures, DEA, consultado el 23 de agosto de 2010 en http://education.qld.gov.au/qcar/pdfs/qcar_white_paper.pdf.

Victorian Curriculum and Assessment Authority (2010), sitio de Internet: www.vcaa.vic.edu.au.

CAPÍTULO 4

Usar los resultados del aprendizaje del alumno para medir las mejoras

4.1 Resultados del aprendizaje del alumno: Instrumentos y medidas de evaluación	51
4.2 El sistema de evaluación ENLACE en México	56
4.3 Desafíos y oportunidades para fortalecer el sistema de evaluación ENLACE	61
4.4 Resumen de recomendaciones para México.....	63

Como se analizó anteriormente, el aprendizaje y el desarrollo del alumno a lo largo del tiempo son criterios claves en los que deben basarse los sistemas educativos, las autoridades educativas locales, las escuelas y los maestros para rendir cuentas. Por lo tanto, un reto importante es evaluar de manera adecuada el aprendizaje y el avance del alumno. Un solo tipo de medida no refleja del todo el aprendizaje del alumno. Todas las formas de medir, desde pruebas estandarizadas hasta portafolios de trabajos del alumno, pueden presentar problemas de validez, fiabilidad y objetividad (Baker, 2010). Es importante elaborar un sistema que emplee no sólo diferentes formas para medir el avance del alumno, sino también diversas fuentes, en las que los resultados de pruebas estandarizadas, por ejemplo, sirvan como referencia cuantitativa (OCDE, 2010).

La primera sección de este capítulo ofrece un panorama de opciones para evaluar el aprendizaje del alumno, y subraya las fortalezas y debilidades de los distintos métodos citando ejemplos internacionales pertinentes. Basada en el análisis de las opciones de evaluaciones formativa y sumativa, la sección sugiere que los sistemas educativos como los de Australia, Alberta (Canadá) y Hong Kong-China se benefician de contar con diferentes fuentes de información, referentes al desempeño del alumno, con el fin de garantizar el mayor nivel de profundidad y exactitud.¹ Sin embargo, contar con una batería de medidas válidas, confiables y variadas del aprendizaje y el progreso del alumno también tiene implicaciones claras en términos de costos y capacidades requeridas, en particular en los ámbitos locales. No sorprende que la mayoría de los países miembros y asociados de la OCDE apliquen evaluaciones a los alumnos basadas en pruebas estandarizadas del aprendizaje del alumno (OCDE, 2008).² Las evaluaciones del aprendizaje del alumno se emplean en varios países con diversos propósitos, por ejemplo, para identificar el desempeño del sistema en su conjunto con fines de diagnóstico vinculados a los esfuerzos de la mejora (como en México), para la rendición de cuentas (como en los Estados Unidos de América y el Reino Unido), para incentivar a los maestros y a las escuelas (como en Chile), y para una combinación de los anteriores (OCDE, 2009a). Como parte integral de un sistema educativo, también las evaluaciones se revisan, evalúan y modifican en los países miembros y asociados de la OCDE para que reflejen de la mejor forma las prioridades de las políticas públicas, las reformas educativas en las áreas concernidas, como son el currículum, y las exigencias de un mundo en constante cambio (como en Australia, Brasil, Noruega, Reino Unido y los Estados Unidos de América, por citar algunos ejemplos).

Aunque no hay un modelo único de evaluación que se destaque en la práctica internacional, algunos de los desafíos técnicos, logísticos y políticos son comunes a todos los sistemas educativos. En este capítulo se presentan algunos de estos problemas comunes y las prácticas recomendadas en términos de vinculación entre las evaluaciones, los estándares y el currículum. Basado en las prioridades de las políticas de un país y en las condiciones, limitantes y oportunidades de un sistema educativo determinado, el reto será identificar la combinación adecuada de diferentes medidas, su importancia relativa, sus usos y las consecuencias (OCDE, 2009b).

Elaborar estos métodos complementarios de medición del aprendizaje de los alumnos requiere tiempo y recursos. Las autoridades educativas pueden establecer un proceso gradual que utilice las fuentes inmediatamente disponibles para las iniciativas de mejora y rendición de cuentas escolares, sin dejar de lado una visión de largo plazo para las evaluaciones. Debido a la relación costo-rendimiento de las evaluaciones estandarizadas de los alumnos, así como a la limitada posibilidad de comparar resultados en diversos contextos nacionales, en varios países miembros y asociados de la OCDE se emplean evaluaciones estandarizadas de aplicación externa, tanto para la rendición de cuentas como para las mejoras. La Evaluación Nacional del Logro Académico en Centros Escolares (ENLACE) en México, que comenzó en 2006, y la *Prova Brasil*, que implementó el gobierno federal brasileño por primera vez en 2005 (recuadro 4.1), son buenos ejemplos de desarrollo dinámico (Zúñiga Molina y Gaviria, 2010; Parandekar *et al.*, 2008). Otros ejemplos en los países miembros de la OCDE muestran que es posible que los sistemas educativos implementen evaluaciones externas, permitiendo al mismo tiempo que los profesionales de la educación innoven en su práctica (OCDE, 2009). Como ejemplo específico de una evaluación que puede propiciar prácticas innovadoras, el capítulo ofrece un panorama de las principales características de la evaluación ENLACE en México, y concluye con consideraciones y recomendaciones específicas para que las autoridades educativas la desarrollen a futuro.

Recuadro 4.1 *Prova Brasil* para la rendición de cuentas y las mejoras

La primera evaluación censal de alumnos en Brasil, *Prova Brasil*, fue aplicada en 2005 por el Ministerio de Educación para medir el logro de alumnos en matemáticas y portugués. La evaluación se aplica cada tres años a alumnos de primaria y secundaria de 4to y 8to grados, y representa una de las principales acciones gubernamentales para establecer un marco de rendición de cuentas orientado a los resultados y enfocado en el avance del alumno. Un estudio reciente realizado por el Ministerio de Educación con apoyo del Banco Mundial empleó los resultados de los alumnos y análisis de la regresión para identificar municipalidades con alto desempeño, incluso después de considerar los antecedentes familiares y socioeconómicos de los alumnos. Con métodos cualitativos, como la observación en el salón de clases y entrevistas, el estudio intentó del mismo modo identificar en el ámbito local las prácticas y políticas adecuadas que contribuían a la obtención de un desempeño superior.

Mayor información (en portugués) se encuentra en <http://provabrasil.inep.gov.br>.

Fuentes: Moriconi, 2009; Parandekar *et al.*, 2008.

4.1 RESULTADOS DEL APRENDIZAJE DEL ALUMNO: INSTRUMENTOS Y MEDIDAS DE EVALUACIÓN

Los resultados de los alumnos, trátense de puntuaciones o de calificaciones reales, o la proporción de alumnos que alcanza niveles de desempeño específicos y predeterminados, pueden tener origen en una o más medidas de aprendizaje del alumno. Estas pueden ser ensayos del alumno, proyectos amplios, portafolios con trabajos del alumno, pruebas de opción múltiple o de respuestas breves, y estas medidas se pueden utilizar para las evaluaciones formativas o sumativas (OCDE, 2009; Baker, 2010).

Es común sugerir que las evaluaciones deben representar de la manera más cercana posible las exigencias cognitivas o habilidades de pensamiento que el sistema educativo considere más importantes. A menudo se prefieren las respuestas abiertas de los alumnos porque obligan a alumnos a recurrir a su propio repertorio de aprendizaje, buscar en él y después aplicar lo aprendido (Baker, 2010). Del mismo modo, para las medidas que abordan las habilidades que se memorizan o desarrollan con facilidad fuera de la escuela (por medio de Internet, por ejemplo), los sistemas educativos pueden elegir procesos de examen más eficientes y rentables, utilizando pruebas extensas y más costosas para el aprendizaje que requiera comprender, aplicar y comunicar de manera elaborada sobre un contenido vasto o complejo (Baker, 2010).

Sin embargo, técnicamente no hay una relación indisoluble entre el nivel cognitivo de la habilidad que se evalúa y el tipo de preguntas que deben emplearse (Zúñiga Molina y Gaviria, 2010). Una consideración importante es la cantidad de información que ofrece una pregunta específica. Si se redacta con cuidado, una sola pregunta de respuesta abierta puede contener más información que la correspondiente pregunta de elección múltiple. Aun así, una serie de preguntas de opción múltiple, previstas para la misma demanda cognitiva, puede proporcionar la misma cantidad de información si son diseñadas cuidadosamente. De este modo se da una compensación entre la información proporcionada en cada pregunta en particular y la facilidad para marcar automáticamente las respuestas dadas por los que se someten a la prueba.³ A menudo, las decisiones referentes a los instrumentos y medidas que van a emplearse se toman con base en el costo y viabilidad, en lugar de hacerlo con base en una medida óptima ligada a estándares específicos (Baker, 2010). Para comparar enfoques, el cuadro 4.1 presenta un resumen de algunas de las principales características, fortalezas, limitaciones, implicaciones de costos y problemas técnicos de distintas medidas.

Cuadro 4.1

Instrumentos y fuentes de evidencia para evaluar el aprendizaje del alumno

Formato	Comentarios/ usos	Fortalezas	Debilidades (fuentes de sesgo, problemas de validez y fiabilidad, consideraciones de costos y requerimientos de capacidad)
Respuestas abiertas y extensas, proyectos, ensayos	Uso formativo, sumativo o combinado. Integrada al currículum; puede requerir asistencia del maestro.	Validez de la actividad, amplio contenido, alta demanda cognitiva; transferencia y aplicación posibles.	Para la rendición de cuentas: costo y posibilidades de comparaciones; costos de reemplazo; formación de maestros para obtener calificaciones válidas y confiables; escasa alineación con estándares.
Portafolios con muestras del trabajo del alumno	Uso en el salón de clases; poca evidencia para la rendición de cuentas, salvo en la selección para programas superiores; puede incluir elementos elegidos del mismo contenido o de distintos.	Medida para identificar desarrollo a lo largo del tiempo, mostrando avance o flexibilidad en múltiples temas; amplias posibilidades para temas, formato o contenido; transferencia y aplicación posibles.	Poca fiabilidad en la calificación si se permite que el alumno elija; posibilidades de comparaciones entre alumnos; costo de la calificación si se realiza fuera del aula; conflicto para los maestros si se emplea para la rendición de cuentas; requiere formación extensa de los maestros; costo elevado si se emplea para la rendición de cuentas.
Observación en aula	Para observar la eficacia del maestro o como oportunidad para explicar resultados obtenidos. Pueden efectuarla personal de la misma escuela, otros maestros, administradores o expertos en pedagogía o en contenido.	Apreciación en tiempo real sobre la actividad de maestros y alumnos; retroalimentación para el maestro; pueden efectuarla otros docentes o el director de la escuela; útil para explicar resultados de alumnos y con modelos de valor agregado; retroalimentación para la evaluación docente.	Requiere acuerdos sobre el modelo de aprendizaje y la relación con estándares; necesita múltiples visitas; observadores capacitados; esquema de clasificación de alta y baja inferencia; si el propósito es la retroalimentación, formación para observar y dar retroalimentación válida de manera confiable; sesgo de observación del comportamiento de maestros y alumnos; no es fácilmente escalable a menos de que se asuman costos significativos de calificación, incluyendo de muestras aleatorias de vídeo.
Pruebas desarrolladas en la escuela y aula, de cualquier formato	Fortalecen la instrucción y alineación; con control de calidad crean un repertorio de actos de evaluación e intervenciones de instrucción; pueden combinarse con evaluaciones sumativas.	Se ajustan al currículum como es enseñado; posibles retroalimentación e intervención inmediatas a alumnos; ofrecen nuevos ejemplos para exámenes a final de ciclo.	Pueden relacionarse poco o mucho con los estándares; pueden ser de calidad deficiente (características psicométricas); los esquemas de calificación pueden no ser explícitos. Se requiere formación en diseño, administración y calificación; si se emplean como parte del sistema de rendición de cuentas, es posible que haya conflicto para los maestros.
Pruebas estandarizadas de opción múltiple y respuestas breves (de procedencia externa)	Utilizadas comúnmente con propósitos amplios de tipo sumativo (de diagnóstico y centrados en la rendición de cuentas).	Calificación de bajo costo; posibilidades para la equiparación vertical y modelos de valor agregado; confiables. El carácter externo reduce el sesgo propio de las relaciones locales.	La validez queda en entredicho si no se ponen a prueba de manera adecuada; el contenido y las demandas cognitivas diseñados cuidadosamente pueden ser superficiales; la alineación con el contenido y con el currículum puede ser débil. Puede fomentar que sólo se enseñe para la prueba y otras conductas no deseables; transferencia limitada de conocimientos y habilidades a escenarios aplicados.

Fuente: Baker, 2010.

Hay una cantidad limitada de enfoques de alta calidad que garantice que se pueden comparar las evaluaciones del desempeño, sobre todo si el interés es brindar algún tipo de retroalimentación sobre el proceso de enseñanza-aprendizaje. Si no se diseñan bien, las evaluaciones quizá no proporcionen la suficiente información sobre los aspectos que se enseñaron bien o mal, lo que daría poca o ninguna orientación para mejorar el sistema. Las evaluaciones del desempeño pueden elaborarse con claras distinciones entre los temas y las necesidades cognitivas y de contenido, en cuyo caso puede ser más fácil la comparación. Cualquier conjunto de limitantes restringe la amplitud de desempeño del alumno que se puede medir. Por lo tanto, se recomienda incluir actividades que no tengan un enfoque específico para ayudar a determinar el grado en que los alumnos transfieren lo que aprenden a situaciones nuevas (Baker, 2010). Por ejemplo, el diseño de las evaluaciones de PISA sigue esta lógica.

El grado de transferencia puede ser relativamente pequeño (pero no por eso difícil), por ejemplo cuando se pide a los alumnos que efectúen una operación de matemáticas presentada en un formato que no hayan visto antes. La transferencia puede ser más difícil cuando se da un problema a los alumnos que requiere la aplicación de varias estrategias, en lugar de un solo procedimiento común para resolverlo (Baker, 2010).

Calificaciones

Por lo general, las respuestas abiertas o conceptuales de los alumnos son calificadas por los maestros, los mismos alumnos o por equipos de maestros con formación especial para calificar exámenes (Baker, 2010). La formación se puede efectuar al hacer que los calificadores examinen diversas respuestas de los alumnos comparándolas con un conjunto de ejemplos ya validados o calificados por expertos. En algunos casos, un jefe de calificadores puede preparar el documento y la calificación se basará en las divergencias con el modelo. En otros casos, la formación implica exponer a los calificadores a la diversidad de formas con que los alumnos pueden alcanzar un cierto nivel de puntuación. En todos los casos, los datos que se capturan se refieren a la eficacia de la formación, por lo general al hacer que el maestro ponga una calificación a una serie de documentos en el nivel que se considere adecuado. Algunas sesiones de formación de calificadores destacan la fiabilidad al centrarse en el grado en que los maestros individuales coinciden entre sí. Este enfoque, cuando se emplea sin validación, puede provocar que los calificadores definan la calidad en términos de las expectativas socialmente compartidas; como un grupo de calificadores puede diferir de otro se aplican distintos niveles de exigencia para distintos grupos de alumnos. Actuar de esta forma socavaría la confianza en los resultados. Con una formación de calidad adecuada, y con rubros comunes de puntuación, la fiabilidad de los calificadores y la validez de sus calificaciones puede ser alta (Baker, 2010).

Logística, costos y tecnología

La formación para calificar exámenes que consisten de trabajos extensos de los alumnos puede requerir entre cuatro horas y dos días (Baker, 2010). A varias sesiones de formación siguen sesiones de calificación, por lo que los costos incluyen el de los calificadores (si son maestros, su tiempo laboral diario normal), viajes, alojamiento y alimentación. En un esfuerzo por limitar los costos, la formación común puede efectuarse en un conjunto de lugares céntricos o a distancia por computadora, con apoyo, obligatorio o disponible, por teléfono o chat. La calificación puede hacerse en casa, con supervisión o sin ella, pero con acceso u obligatoriedad de llamadas telefónicas del supervisor de calificaciones. En algunos contextos educativos, la calificación de exámenes se asigna a terceros, maestros o no, en otros locales, pero esto no es algo común (Baker, 2010). Los individuos pueden recibir la formación en persona o a distancia. Cuando la calificación se efectúa sin supervisión (por ejemplo, en casa del maestro o calificador), los costos de la calificación se reducen de manera significativa, dado que no se requieren gastos de viaje ni otros viáticos, y el pago se puede hacer dependiendo de la cantidad de documentos calificados. En algunos países y jurisdicciones regionales, la calificación de documentos es algo que por rutina se espera que hagan los maestros, por lo que se incluye en toda negociación colectiva con sus sindicatos (*i.e.* Alberta, Canadá), mientras que en otros, la calificación de evaluaciones “no es necesario que la hagan los maestros”.⁴

En fechas más recientes se ha empleado la tecnología para calificar ensayos de alumnos con un grado razonable de confiabilidad. Los enfoques implican el uso de ensayos precalificados y un complejo modelo de regresión que incluye aspectos lingüísticos y léxicos del trabajo de los alumnos (Burstein, 2003). Algunos de estos enfoques requieren una formación que necesita tiempo y la clasificación de documentos antes de que la computadora proceda a la calificación, procedimiento que debe efectuarse por cada cambio de tema. Otros enfoques funcionan de manera semejante a los verificadores de gramática y ortografía de programas de procesadores de texto. También existen complejos sistemas que comprenden el idioma natural, aunque hasta la fecha siguen requiriendo un trabajo intensivo para adaptarlos a distintos temas y niveles de trabajo de los alumnos (Baker, 2010). La puntuación por computadora de temas como la resolución de problemas y otras

respuestas abiertas está en desarrollo (Chung y Baker, 2003; Chung *et al.*, 2001), y es más útil si la pregunta tiene un conjunto específico de respuestas correctas (Baker, 2010). La puntuación por computadora también rastrea (mediante redes neurales) los patrones adoptados por distintos alumnos para contestar correctamente. Estos datos son útiles para evaluaciones formativas (Baker, 2010).

Sin embargo, dos de las principales limitaciones para la calificación por computadora han sido la disponibilidad del equipo y el conocimiento sobre computadoras de alumnos y maestros. Ha habido mejoras significativas en el programa de barrido óptico de la escritura y reconocimiento de voz de los alumnos. En unos cuantos años esto puede tener repercusiones importantes para la práctica actual de la evaluación en papel a nivel mundial (Baker, 2010). Sin embargo, las limitaciones estructurales para acceder a las computadoras, conocer su manejo y conectividad aún son desafíos para las economías emergentes y en vías de desarrollo (ITU, 2009).⁵

Calidad de las evaluaciones

La calidad técnica de las evaluaciones es un aspecto importante cuando se emplean sus conclusiones para tomar decisiones importantes para alumnos, maestros, directores u otras personas.⁶ Un criterio clave de la calidad técnica de un examen es la *validez*. La validez depende del propósito de la prueba y la evidencia de que los usos de la prueba son adecuados. Si el propósito de la prueba es evaluar con precisión la adquisición de un contenido y de ciertas habilidades por parte de los alumnos, la inspección de las actividades o preguntas y la profundidad estimada de muestreo son consideraciones importantes para verificar la validez del contenido. Las antiguas nociones de validez, como la validez “aparente”, o de contenido, concurrente y predictiva, fueron reunidas en una consideración global de la validez de las pruebas o exámenes (Baker, 2010; Linn [ed.], 1993).

Si el propósito de la prueba es seleccionar candidatos para niveles más elevados de escolaridad, entonces puede examinarse la prueba respecto a su capacidad de predecir el éxito de los alumnos en una escolaridad más avanzada. Como los sistemas de rendición de cuentas y sus evaluaciones han evolucionado, las evaluaciones parecen tener propósitos múltiples, lo cual aumenta la dificultad de estimar la validez. Por ejemplo, si se diseña una prueba de rendición de cuentas para situar con precisión a los alumnos dentro de una clasificación o línea continua que refleje su nivel en el dominio de las materias, la evaluación necesita ser capaz de diferenciar entre alumnos de distintos niveles clases mediante preguntas más difíciles para aquellos que se espera tengan habilidades más desarrolladas. Si se pretende que la prueba mida la eficacia del sistema educativo, se debe disponer de una evidencia de validez que demuestre que la prueba es útil para una enseñanza de alta calidad. Si el desempeño de la prueba no cambia en función de la enseñanza, sino en función de la madurez o de otras influencias ajenas a la escuela, entonces no es adecuado usarla en sistemas de rendición de cuentas. Además, tal vez se espere que los maestros revisen su periodización de la enseñanza con los resultados del mismo examen, ya sea en el mismo año lectivo o en otros años. Por tanto, debe haber evidencias de que los resultados reportados ofrecen la suficiente información pertinente para dicha función. Si se espera que las pruebas supervisen el crecimiento de los alumnos durante varios años, es esencial la idea de la equiparación vertical (la dificultad de las pruebas y las escalas son equivalentes en distintos años). Este requerimiento también sirve de apoyo a los modelos de valor agregado que pretenden identificar los aportes de las escuelas a la mejora de los resultados del alumno (esto se trata en mayor detalle en el capítulo 5).

Una revisión de los resultados de las pruebas debe considerar si el examen es pertinente para sus propósitos, y si hay evidencias que documenten la capacidad del examen para cumplir con sus propósitos o aspiraciones. Es difícil obtener evidencia de validez, en especial durante un periodo de pruebas de desarrollo, cuando no existe la perspectiva de un reporte o sanciones públicos. Puede experimentarse para determinar si las evaluaciones son sensibles a la instrucción, si categorizan de manera adecuada a los alumnos que se desempeñaron bien en medidas semejantes (validez de concurrencia) o si los estándares de desempeño pertinentes se fijaron en un nivel apropiado (en contraste con niveles definidos políticamente) (Baker, 2010).

Recuadro 4.2 Sistemas mixtos de evaluaciones de alumnos

Victoria, Australia: combinar evaluaciones a nivel de escuela y estatales

En el estado de Victoria, Australia, la Autoridad sobre el Currículo y la Evaluación de Victoria (*Victorian Curriculum and Assessment Authority*, VCAA por sus siglas en inglés) estableció los Estándares de Aprendizaje Esencial en Victoria (*Victoria Essential Learning Standards*) para brindar una descripción anual de lo que se espera que aprendan y realicen todos los alumnos de primaria y secundaria. La VCAA también maneja el Programa Nacional de Evaluación – Lectoescritura y Pensamiento Matemático (*National Assessment Program – Literacy and Numeracy*, NAPLAN por sus siglas en inglés) y proporciona a las escuelas evaluaciones en la escuela a petición suya. Los maestros participan en la elaboración de las evaluaciones desarrolladas en la escuela junto con el personal de apoyo académico, y todas las evaluaciones de años anteriores están disponibles al público. Al menos 50% de la puntuación total de los alumnos consiste en actividades propias del aula (por ejemplo, experimentos en el laboratorio, la investigación sobre temas clave y reportes extensos). Al mismo tiempo, como parte del NAPLAN, aproximadamente 260 000 alumnos de los grados 3, 5, 7 y 9 se someten a evaluaciones estandarizadas en toda Australia. Así, el sistema combina evaluaciones basadas en la escuela con evaluaciones estandarizadas estatales.

Para mayor información: www.vcaa.vic.edu.au

Alberta, Canadá: Elaboración de un marco integral para la evaluación

En Canadá, la Autoridad para la Educación de Alberta (*Alberta Education Authority*) encargó el Estudio de la Evaluación de los Alumnos en Alberta (*Alberta Student Assessment Study*), una revisión sobre la teoría y las prácticas referentes a las evaluaciones de los alumnos que generó recomendaciones sobre:

- Los resultados del aprendizaje curricular, estándares de desempeño y reportes de logro del alumno.
- La manera como las evaluaciones externas y las basadas en el aula pueden ser usadas de manera óptima para brindar información al tomar decisiones importantes sobre las necesidades de los alumnos, el manejo de la escuela y los problemas enfrentando a la provincia relativos a garantizar la oportunidad de aprendizaje para todos los alumnos.

Los resultados del estudio se presentaron en 2009, con directrices y recomendaciones específicas sobre la manera en que el sistema educativo puede combinar eficazmente estándares de desempeño, evaluaciones en el aula y evaluaciones de la provincia, reportes de logros del alumno y desarrollo profesional de los maestros. Sus recomendaciones se basan en evidencias sólidas, y brindan referencias útiles para otros sistemas que buscan crear enfoques complementarios para evaluar a los alumnos dentro de un marco de rendición de cuentas y de mejoras.

Para mayor información: <http://education.alberta.ca/department/ipr.aspx>

Fuentes: Victorian Curriculum and Assessment Authority, 2010; Darling-Hammond y McCloskey, 2008; Government of Alberta (Canada) – Education, 2009.

Del mismo modo la *fiabilidad* de las evaluaciones es un asunto vital. Esto se refiere a la consistencia con que una prueba mide el desempeño. La fiabilidad puede degradarse de manera natural si se tiene la intención de cambiarla (por ejemplo, si se pretende mejorar). Este problema se puede resolver en parte con controles estadísticos adecuados. Las medidas empleadas en la rendición de cuentas también deben vencer el reto de la *equidad*. Equidad no significa resultados iguales, sino que las características del examen y la calificación no favorezcan a ningún grupo particular, salvo a los que estén mejor preparados. **Para que los sistemas educativos que atienden a grupos heterogéneos de alumnos (de diversos orígenes socioeconómicos, étnicos o con una lengua materna distinta),**

la equidad en la evaluación es un asunto importante. Las características lingüísticas de los alumnos pueden confundir las estimaciones del aprendizaje en otras materias, como en matemáticas o en ciencias. Otros problemas son que las actividades sean pertinentes sólo para un subconjunto de alumnos (por ejemplo, urbano o rural) o muestren diferencias de desempeño independientes de la competencia en otras actividades semejantes en el área. Entrelazar estos requerimientos técnicos en una tela que soporte la calidad técnica de las medidas empleadas en la rendición de cuentas, requiere previsión y análisis por parte de los actores implicados para determinar las características que puedan plantear desafíos para una aceptación amplia (Baker, 2010).

Como todas las mediciones del aprendizaje del alumno pueden presentar desventajas potenciales, fuentes de error y parcialidad que afecten la validez y la fiabilidad de los resultados, los sistemas educativos pueden optar por tener diversas fuentes de información sobre el desempeño del alumno para garantizar el mayor nivel de validez y exactitud (Baker, 2010). Las evaluaciones del aprendizaje del alumno se emplean en diferentes países con diversos propósitos. Por ende, el reto es encontrar el equilibrio adecuado entre evaluaciones estandarizadas y evaluaciones basadas en la escuela, calificadas y referenciadas de manera externa e interna, para ser usadas con diferentes fines y de acuerdo con las limitaciones estructurales, presupuestales y de capacidad del sistema educativo. La siguiente sección revisa una importante evaluación de alumnos en México, y analiza algunas de sus principales características a la luz del análisis previo, con el fin de identificar desafíos y oportunidades para su desarrollo a futuro.

4.2 EL SISTEMA DE EVALUACIÓN ENLACE EN MÉXICO

En 2006, la SEP implementó por primera vez la Evaluación Nacional del Logro Académico en Centros Escolares (ENLACE). La prueba ENLACE se diseñó para brindar información a alumnos, padres de familia, maestros, directores y al público en general sobre el logro individual de los alumnos, agrupando los resultados por escuelas.⁷ En contraste con los exámenes de EXCALE, que se administran sólo a muestras de alumnos en distintos grados,⁸ el propósito original de la prueba ENLACE fue servir como punto de referencia para mostrar mejoras en los procesos de enseñanza y aprendizaje, en las aulas y escuelas, a nivel primaria y secundaria (Zúñiga Molina y Gaviria, 2010).

Desde 2006 se incluyeron las materias de matemáticas y español en cada sesión, y una tercera materia que cambia cada año: ciencias se incluyó en 2008, formación cívica y ética en 2009, historia en 2010; en 2011 se incluirá geografía. En la actualidad, el examen se aplica a alumnos de primaria de tercer a sexto grado, y a los de secundaria, de primero a tercer grado. En 2009 la prueba se aplicó por primera vez a los dos primeros años de secundaria. La prueba ENLACE se aplicó en 2009 a más de la mitad (51%) de todos los alumnos de los niveles de preescolar, primaria y secundaria en México (INEE, 2010, indicador ED01; Zúñiga Molina y Gaviria, 2010).

Los niveles del avance de los alumnos reflejados en los resultados de la evaluación ENLACE desde 2006 coinciden en general con los resultados de 2006 de PISA: ambos presentan un promedio bajo, pero al mismo tiempo hubo una mejoría. Por ejemplo, el porcentaje de alumnos clasificados entre 2006 y 2009 como “insatisfactorio” o “regular” cayó de 78.7% a 67.2%, y el porcentaje de alumnos clasificados como “bueno” o “excelente” subió de 21.3% a 32.8%. Los resultados de la aplicación de la prueba ENLACE en 2010, así como los resultados de 2009 de PISA, brindarán más información sobre las tendencias del avance de los alumnos.

Administrada por la SEP por conducto de la Dirección General de Evaluación de Políticas (DGEP), la evaluación ENLACE es ya una medición del desempeño del alumno socialmente aceptada (Zúñiga Molina y Gaviria, 2010; Salieri *et al.*, 2010). La prueba se aplica en abril de cada año y los resultados se publican en septiembre del siguiente año lectivo. La DGEP procesa los resultados de la prueba utilizando programas de computación tanto comerciales como propios, y genera materiales para que los usuarios interpreten los resultados, difundidos en paquetes de información escolar y por Internet (www.dgep.sep.gob.mx). En esta página electrónica, alumnos, familias y toda persona interesada pueden obtener información mediante un número especial de identificación presente en la hoja de respuestas del alumno. Los resultados también se encuentran en forma agrupada por escuela,⁹ por estado o a nivel nacional. La cobertura de los resultados por parte de los medios es muy amplia,

y aunque los funcionarios desalientan esta práctica, son comunes diversas versiones de “cuadros escolares” que comparan promedios agrupados de calificaciones brutas de los alumnos por escuela. La importancia que la SEP y las autoridades educativas estatales dan a la publicación de resultados de ENLACE se sustenta en comparaciones internacionales. Como se analizó en el capítulo 3, los resultados de 2006 de PISA indican que un impacto significativo en el desempeño del alumno, en todos los países, se relaciona con la publicación de los datos del logro de alumnos por escuela (OCDE, 2007).

La información que se supone reciben las escuelas por conducto de las autoridades educativas estatales incluye la proporción de alumnos en cada nivel de avance por grado y contenido de las materias. Cada escuela debe recibir también información sobre la proporción de alumnos en cada nivel de avance en comparación con los resultados de los alumnos y escuelas del mismo tipo, en los ámbitos estatales y nacional. La información se organiza de modo que sirva para identificar las posibles oportunidades de mejorar la enseñanza, y para permitir que los grupos comparen sus resultados con los de otras escuelas en condiciones socioeconómicas y de infraestructura semejantes (Zúñiga Molina y Gaviria, 2010).¹⁰ Así, maestros, directores de escuela, y alumnos y sus familias pueden evaluar el progreso y las dificultades que encontraron en el aprendizaje, además de identificar las partes del currículum que no se cubrieron de manera adecuada. Se espera también que los maestros analicen los resultados de los alumnos e identifiquen fortalezas y debilidades en las áreas temáticas de la prueba.

Una revisión reciente de los usos de ENLACE en el ámbito estatal reveló que la evaluación se ha convertido en una referencia nacional y local sobre el avance del aprendizaje de los alumnos, y que la mayoría de las autoridades educativas estatales efectúan algún tipo de actividad de seguimiento (Salieri *et al.*, 2010). La SEP proporciona a todas las autoridades educativas estatales folletos impresos, informes y CDs sobre el desempeño individual y agrupado por escuelas para que se distribuyan en las escuelas. Algunos estados como Jalisco, Nuevo León y Veracruz elaboran su propio material, que se distribuye al personal supervisor (Supervisores o Jefes de sector), y ofrecen algún tipo de apoyo y cursos de desarrollo profesional a las escuelas identificadas como de bajo desempeño, según los resultados generales de la prueba ENLACE y con base en sus necesidades (Salieri *et al.*, 2010). Esto pone de relieve la importancia y las oportunidades de las autoridades educativas estatales respecto a las acciones de mejora y los mecanismos de rendición de cuentas (que se tratan en el capítulo 7 y otros de este informe).

Diseño y características técnicas

Basado en un análisis de los elementos básicos de diseño, la características y los resultados de las pruebas desde 2006, el instrumento de evaluación ENLACE presenta niveles sólidos de consistencia, validez y fiabilidad internas como medición del aprendizaje del alumno (Zúñiga Molina y Gaviria, 2010).¹¹ Sin embargo, es importante observar que el desarrollo de la prueba ENLACE requiere tal vez estudios más profundos, en particular a la luz de las reformas en curso del currículum (contenidos) en México, así como los usos que se puedan asignar a la evaluación ENLACE en un futuro cercano. A continuación se resumen los elementos de la evaluación que fueron revisados, con conclusiones preliminares respecto a su validez y fiabilidad.

Modelo psicométrico

Las listas de especificaciones para ENLACE fueron establecidas en un principio por expertos en currículum de la misma SEP, quienes determinaron cuál era el contenido más pertinente para cada prueba. El personal de la DGEP afinó y amplió las listas con ayuda de expertos del INEE y la aprobación de la Subsecretaría de Educación Básica de la SEP. Al elaborar la evaluación se establecieron tres niveles de dificultad (baja, media y alta) con el fin de garantizar que las listas describiesen el contenido por evaluar en cada grado, el nivel de dificultad y la materia. Las especificaciones para elaborar las preguntas de la prueba están disponibles al público y están abiertas a revisiones a partir de sugerencias de maestros, directores y autoridades educativas estatales. Sin embargo, las listas para matemáticas y español no han cambiado desde 2006. Las listas admiten interpretaciones cualitativas de las diferencias en el desempeño entre alumnos con el fin de permitir que la retroalimentación mejore los procesos basados en los resultados. La escala con que se reportan los resultados de la prueba ENLACE tiene una media

de 500 puntos, con una desviación estándar de 100 puntos correspondiente a los promedios de 2006 que sirven como primer punto de referencia. Los resultados de la evaluación ENLACE tienen una distribución normal, y según ésta el 99% de los alumnos obtiene una puntuación de entre 200 y 800 puntos por grado y materia. La escala se basa en la teoría de respuestas por reactivo (*Item Response Theory*, IRT por sus siglas en inglés) y asigna a los alumnos distintos niveles de avance, permitiendo la comparación entre años sucesivos. Las preguntas de la prueba se analizan antes de la puntuación mediante un modelo clásico (índice de dificultad y correlación de dos series de puntos, como aproximación por distinción de preguntas). Después se ajustan las preguntas y se clasifica a los alumnos de acuerdo con los tres parámetros del modelo IRT. A cada alumno se le asigna un valor de puntuación considerando no sólo la cantidad de respuestas correctas, sino también qué preguntas contestó correctamente. Como se fija una escala por cada grado y por el contenido de cada materia del currículum, no es posible comparar las puntuaciones entre diferentes grados o años lectivos.

A diferencia del modelo Rasch, en el modelo de tres parámetros los conceptos por medir se definen antes de adoptar el modelo de medición. Para la elaboración de ENLACE, el currículum vigente orientó la creación y selección de preguntas para la prueba, y los parámetros se ajustaron a las características de las preguntas. En el modelo Rasch, la medición de la invariancia no se puede obtener al aplicar únicamente el modelo a una serie determinada de preguntas; el modelo psicométrico es el principio que define al concepto, en lugar de que sea el concepto el que dirija la elaboración de la prueba. El modelo de tres parámetros usado por ENLACE permite que la prueba refleje la estructura del currículum sin poner en riesgo el modelo de medición. Debe observarse que los puntajes de corte para los niveles de logro no son los mismos en todos los grados, ya que fueron definidos por separado para cada grado y materia (es decir, no hay una escala común para todos los grados).

Dimensionalidad

La dimensionalidad de las pruebas ENLACE es una de las características fundamentales que deben analizarse para determinar la estabilidad estructural de los resultados de la evaluación en el tiempo y, por ende, permitir estimaciones de mejoras reales en el aprendizaje alcanzado por los alumnos. Un estudio reciente de ENLACE a cargo de Hernández y Olariaga (2009) concluyó que:¹²

- Con pocas excepciones, las pruebas de ENLACE aplicadas en el año lectivo 2008-2009 esencialmente pueden considerarse unidimensionales o con una multidimensionalidad baja.
- Los resultados de muestras tomadas de la población y de control no sugieren estructuras dimensionales distintas.
- En general se puede considerar que las pruebas tienen una complejidad baja o una estructura sencilla.
- Respecto a las posibles diferencias de dimensionalidad en las pruebas, hay cierto nivel de multidimensionalidad en grados específicos. Esto quizá se deba a la mayor complejidad del contenido del currículum en esos grados (en el tercer año de secundaria).¹³

Estos resultados sugieren que las características de la evaluación ENLACE, combinadas con la solidez de los modelos de creación de las pruebas, garantizan la equiparación correcta de las respuestas de los alumnos.

Fiabilidad

Basándose en el coeficiente de consistencia interno calculado (coeficiente alfa de Cronbach), la fiabilidad de las pruebas de ENLACE es alta. Los valores calculados para ENLACE se ubican dentro del rango generalmente aceptado como indicador de puntuaciones bastante confiables. Por ejemplo, para las pruebas de ENLACE de 2006 a 2008, los valores de coeficiente alfa varían entre 0.75 y 0.92, con promedios en español, matemáticas y ciencias muy por arriba de 0.80 (cuadro 4.2). Basada en una comparación de valores previstos para PISA del 2000 (OCDE, 2002, p. 152, cuadro 4.1), la fiabilidad de la prueba ENLACE es semejante, y en algunos casos rebasa la de los resultados de PISA del 2000 (escala unidimensional incondicionada).

Cuadro 4.2
Fiabilidad de ENLACE

Materia	Año	Fiabilidad de la evaluación		
		2006	2007	2008
Matemáticas	3o de primaria	0.896	0.896	0.912
	4o de primaria	0.915	0.898	0.922
	5o de primaria	0.896	0.874	0.907
	6o de primaria	0.872	0.874	0.910
	3o de secundaria	0.838	0.789	0.865
Español	3o de primaria	0.876	0.844	0.879
	4o de primaria	0.903	0.900	0.906
	5o de primaria	0.809	0.837	0.804
	6o de primaria	0.880	0.891	0.910
	3o de secundaria	0.835	0.813	0.752
Ciencias	3o de primaria			0.854
	4o de primaria			0.853
	5o de primaria			0.818
	6o de primaria			0.880
	3o de secundaria			0.804

Fuente: Zúñiga Molina y Gaviria, 2010.

Validez

Se efectuó un estudio para evaluar simultáneamente la validez de ENLACE en relación con otras pruebas, como la de PISA. Como no se contaba con suficientes elementos de PISA accesibles públicamente, los investigadores crearon una prueba especial (SEP-ISA) en convenio con el Consejo Australiano para la Investigación Educativa (*Australian Council for Educational Research*), utilizando preguntas del banco de preguntas con que se elaboró el PISA, así como con preguntas ya empleadas en las pruebas de PISA y difundidas posteriormente. Los investigadores seleccionaron para este estudio una muestra aleatoria estratificada de 11 717 alumnos de segundo y tercer años de secundaria de todo México. Estos alumnos realizaron tanto las pruebas de matemáticas y español de ENLACE, como la prueba SEP-ISA (matemáticas y comprensión de lectura).¹⁴ La correlación de las escalas de las distintas pruebas, corregidas por atenuación, fue aproximadamente de 0.829 para español/comprensión de lectura y de 0.810 para matemáticas (Zúñiga Molina y Gaviria, 2010). En el cuadro 4.3 se presentan con fines comparativos las correlaciones estimadas entre escalas y sub-escalas de matemáticas, lectura y ciencias de PISA de 2003. Las correlaciones obtenidas entre ENLACE y SEP-ISA son de la misma magnitud de las estimadas para la sub-escala de resolución de problemas con las demás dimensiones de matemáticas en PISA, que son de 0.79 a 0.83. Comparativamente, estos resultados sugieren que los niveles de validez de ENLACE son muy altos.

Calidad del proceso de equiparación

Como se ha descrito antes, el contenido de aprendizaje específico que evalúa la prueba ENLACE se determina por las listas de especificaciones. Para identificar, jerarquizar y enfocar el contenido del currículum para las pruebas de estas listas se utilizaron cuatro criterios: *pertinencia*, *credibilidad*, *continuidad* y *amplitud*.¹⁵ Cada año deben crearse nuevas versiones de las pruebas de ENLACE porque los cuadernillos de las pruebas se divulgan al público después de su aplicación. Quienes elaboran las pruebas emplean una variante del diseño de población común para garantizar que estas sean equivalentes en años consecutivos para el mismo contenido de la materia y grado.

Cuadro 4.3

Correlación entre sub-escalas de resolución de problemas, lectura y ciencias, PISA 2003

	Espacio y forma	Cambio y relaciones	Incertidumbre	Cantidad
Espacio y forma		0.89	0.88	0.89
Cambio y relaciones			0.92	0.92
Incertidumbre				0.9
Resolución de problemas	0.79	0.83	0.81	0.82
Lectura	0.67	0.73	0.73	0.73
Ciencias	0.73	0.77	0.77	0.76

Fuente: OCDE, 2005a, p. 190, cuadro 13.4.

La adecuación del proceso de equiparación depende de la estabilidad de las estimaciones del parámetro de las preguntas y de las puntuaciones de los alumnos: las estimaciones de los parámetros deben ser consistentes, sin importar el subconjunto de preguntas de la estimación. Para evaluar ENLACE en este rubro, Zúñiga Molina y Gaviria (2010) estimaron los parámetros de las preguntas cuando se ajustaron las preguntas por separado y cuando estas mismas preguntas se ajustaron junto con las preguntas de la “forma previa a la prueba”.¹⁶ Por cada año y área temática se obtuvieron para las estimaciones del parámetro de la pregunta dos variables correspondientes a la dificultad. Se observó que estas dos variables tenían una correlación alta en todos los grados y materias, pues nunca descendieron de 0.993 (Zúñiga Molina y Gaviria, 2010, apéndice I). Los autores también compararon las puntuaciones de los alumnos en un nivel de año (grado), de nuevo con dos variables por grado y materia. También se observó que estos valores tenían una correlación elevada, nunca menor a 0.985 (Zúñiga Molina y Gaviria, 2010, apéndice II).

Asimismo se calculó la equiparación de errores en las pruebas de ENLACE de 2008 y de 2009, comparándola después con los datos de equiparación de errores para el componente de comprensión de lectura de PISA de 2003. Se descubrió que los valores de la equiparación de errores fueron muy similares entre los resultados de 2008 y 2009 de ENLACE, y en algunos casos menores a los valores de equiparación de errores, como en la comprensión de lectura, de PISA de 2003. Esto sugiere que para ENLACE el proceso de equiparación horizontal es confiable pues permite comparar los resultados del alumno por cada grado y materia en años consecutivos (es decir, distintos alumnos en el mismo grado) (Zúñiga Molina y Gaviria, 2010, p. 39).

Equiparación vertical

Aunque la evaluación ENLACE no está diseñada para una equiparación vertical, la SEP y expertos invitados efectuaron estudios de viabilidad para determinar las opciones para fortalecer la prueba ENLACE en un futuro cercano e incluir una escala vertical. Por ejemplo, esto permitiría comparar los resultados de alumnos entre distintos grados (es decir, potencialmente los mismos alumnos en diferentes años). Los estudios preliminares se centraron en los resultados de 104 487 alumnos en el componente de matemáticas de ENLACE del sexto año de primaria (4 533 aulas) en la Ciudad de México (presentado en el apéndice III de Zúñiga Molina y Gaviria, 2010). Los resultados de estos ensayos muestran que es posible que ENLACE incluya una escala común entre quinto y sexto grados de matemáticas. Más aun, la caída en los resultados en casi un 36% en alumnos de entre quinto y sexto años, se asemeja a los resultados de los alumnos de 15 años de edad evaluados en PISA. La equiparación vertical también permitiría poner a prueba las puntuaciones de corte definidas para cada nivel de grado en ENLACE. Por ejemplo, los puntos de corte para todos los años de escuela primaria incluidos en ENLACE (de tercero a sexto), deberían revisarse basándose en una reconsideración de los criterios con que se establecieron, así como en la validez de la escala vertical. Por último, también se necesitaría estudiar el margen de error de la equiparación vertical antes de incorporar una escala vertical común en la prueba ENLACE. El fortalecimiento posterior de ENLACE debe incorporar estas consideraciones, así como otras que se bosquejan en la sección 4.3.

Cuadro 4.4

Porcentaje de trampa probable detectado en ENLACE de 2006 a 2009

Año	2006	2007	2008	2009
3o de primaria	6.28	5.50	10.16	6.97
4o de primaria	4.90	5.60	7.53	6.65
5o de primaria	4.14	3.68	4.78	4.45
6o de primaria	4.10	4.86	5.48	4.75
1o de secundaria				1.54
2o de secundaria				3.79
3o de secundaria	3.24	3.14	7.08	6.54

Fuente: Zúñiga Molina y Gaviria, 2010.

El factor copia

Como ENLACE es una evaluación de censo (es decir, se pretende evaluar a todos los alumnos del grado pertinente), la supervisión y control de las condiciones de la prueba y la aplicación de la prueba se presentan como desafíos, en particular por la gran diversidad de contextos escolares entre y al interior de los estados. Para identificar la magnitud de la probable copia de respuestas durante los exámenes, se emplean dos métodos. Si bien los porcentajes promedio de las irregularidades (“trampa”) probables ascendieron hasta un 7% en 2008 en comparación con el 4.5% de 2006, la tendencia se redujo en 2009 con un promedio de 6.5% (cuadro 4.4). Los resultados preliminares de la sesión de 2010 de ENLACE confirman que la copia de respuestas no aumentó. Además, el análisis de los efectos generales de copiar, a cargo de Zúñiga Molina y Gaviria (2010), muestra que incluso en 2008 las estimaciones de validez y fiabilidad de ENLACE en gran medida se mantuvieron intactas.

Sin embargo, debe mantenerse la supervisión y el análisis del factor de copia de respuestas, y tomar medidas para resolverlo en cada aplicación. Asimismo deben incluirse recursos adicionales y mecanismos de supervisión en la planeación a futuro de la evaluación ENLACE.

4.3 DESAFÍOS Y OPORTUNIDADES PARA FORTALECER EL SISTEMA DE EVALUACIÓN ENLACE

En México, como en otros sistemas educativos de mejor desempeño, existe cada vez más un mayor consenso sobre los beneficios de definir con claridad el avance que se espera alcancen los alumnos en la adquisición de habilidades y competencias. Por lo tanto, en años recientes la SEP emprendió reformas del currículum enfocadas en desarrollar las habilidades de los alumnos, poniendo un acento importante en los resultados específicos del aprendizaje. Estas reformas han avanzado de manera significativa en los niveles de preescolar y secundaria, y se encuentran en una fase experimental en el nivel primaria. En la medida en que estas reformas se enfoquen en el desarrollo de competencias y habilidades para la vida, ENLACE necesitará reflejar esos cambios. Con un marco basado en estándares en México, las pruebas referentes al currículum necesitarán evolucionar para reflejar los estándares de competencias y habilidades que se quieren establecer. Estándares de contenido y desempeño bien definidos para los alumnos, elaborados como parte de las reformas del currículum en México, servirían como punto de partida y referente para la planeación docente, los materiales educativos, las prácticas de la enseñanza, la capacitación, el desarrollo profesional y, en última instancia, las evaluaciones. Quizá lo más importante sea contar con estándares claros para el aprendizaje y el desarrollo del alumno, pues esto proporcionaría una visión coherente y generaría expectativas comunes que compartieran todos los niveles y actores educativos, incluyendo a las autoridades educativas estatales, los padres de familia, los comités y consejos escolares, y por los que podrían trabajar de manera conjunta. En este contexto, deben considerarse los siguientes puntos y recomendaciones:

- **La necesidad de conservar los niveles de fiabilidad, validez, estabilidad y sencillez estructurales que ya alcanzó ENLACE**, sin dejar de tomar en cuenta las transformaciones del currículum que están en análisis y que puedan implementarse.

- **La necesidad de establecer un programa de desarrollo claro para ENLACE en el que se definan los objetivos, los fines y los tiempos de las políticas educativas en torno a la prueba.** El mejoramiento posterior de ENLACE debe incluir estudios pertinentes para determinar el grado de comparación vertical entre alumnos y grupos con el fin de medir el avance hacia expectativas definidas del aprendizaje. Las mediciones del progreso del alumno deben separarse de los factores socioeconómicos y otros aspectos pertinentes para identificar el aporte real de las escuelas, zonas escolares, regiones y estados a los resultados de los alumnos.
- **La necesidad de abordar aspectos administrativos, técnicos y logísticos, para obtener con el paso del tiempo medidas más confiables del avance de alumnos hacia objetivos específicos de aprendizaje** (tal como son los estándares de contenido del aprendizaje del alumno). Abordar estos elementos no sólo fortalecería a ENLACE, sino que también contribuiría a robustecer el marco de evaluación en México para la rendición de cuentas y las acciones de mejora (capítulo 3). Más aun, abordar los siguientes puntos permitiría elaborar los métodos de valor agregado que se presentan con más detalle en el capítulo 5 de este informe y en su publicación hermana *La medición del aprendizaje de los alumnos: Mejores prácticas para evaluar el valor agregado de las escuelas* (OCDE, 2011).

Condiciones administrativas

- **Integridad y consistencia en la identificación de alumnos, maestros, escuelas y directores cuando sea adecuado.** Debe haber mecanismos específicos para detectar e incorporar a los individuos que tal vez no se encuentren en las bases de datos y para corregir inconsistencias. La movilidad de alumnos y maestros entre y al interior de las zonas escolares y estados debe ser identificable con un seguimiento adecuado.
- **Expedientes individuales unificados que acompañen a cada alumno a lo largo de toda su vida escolar.** Debe incluirse información sobre los resultados de las evaluaciones, incluyendo ENLACE, para determinar el progreso en el aprendizaje. Este expediente serviría asimismo para los alumnos vulnerables y para las acciones tendientes a reducir los índices de abandono escolar.
- **Referencias estándar y únicas para ciudades, pueblos, municipios y escuelas.**
- **Capacidad para asociar y combinar el logro de los alumnos en la prueba ENLACE con los maestros que les brindaron clases.**
- **Capacidad para hacer que corresponda cada punto en la base de datos de los alumnos con sus contrapartes en las bases de datos de los maestros, directores y escuelas,** con el fin de determinar la contribución de distintos maestros, o de toda la escuela, a las ganancias de aprendizaje de cada alumno.

Condiciones logísticas

- **Una mejora sustancial en las condiciones de la aplicación de la prueba.** Esto implica contar con mecanismos para limitar conductas no deseadas por parte de los maestros y directores, así como abordar el problema del copiado de las respuestas. Las necesidades en cuanto a la seguridad, la supervisión y las condiciones adecuadas y estandarizadas para aplicar la prueba ENLACE irán necesariamente en aumento, por lo que la SEP y las autoridades educativas estatales deberán de considerar estos costos en su planeación.

Condiciones técnicas

- **Robustecimiento técnico continuo de las evaluaciones de ENLACE.** La validez, consistencia interna y fiabilidad demostradas de ENLACE sugieren que sólo cambios sustanciales en el currículum deberían de implicar modificaciones correspondientes en la prueba. Sin embargo, en vista de las reformas del currículum que se analizan o se implementan actualmente en México, esto puede ofrecer una oportunidad para planear el desarrollo de ENLACE que asegure que la alineación, la coherencia, la demanda y amplitud cognitivas, correspondan con los objetivos de las políticas educativas en México.

- **Implementación de un diseño de equiparación vertical para todos los grados y contenido de materias que permita calcular el progreso educativo por alumno.** La SEP y los investigadores invitados necesitarán calcular los errores en la equiparación para comparar datos, verificar en qué medida los errores son aditivos en el mismo grupo y en grados sucesivos, y si la magnitud de estos errores acumulativos impide, por ejemplo, que la escala avance más de dos o tres años. Los puntos de corte deben ser definidos de nuevo para que sean consistentes en grados consecutivos.

4.4 RESUMEN DE RECOMENDACIONES PARA MÉXICO

Basado en las consideraciones anteriores, a continuación se presenta un resumen de las recomendaciones para México sobre la importancia de evaluar los resultados del aprendizaje del alumno, las oportunidades que ofrece ENLACE, y los retos y oportunidades para su desarrollo a futuro:

- **El aprendizaje y el avance del alumno como base de la rendición de cuentas y de los estándares requieren múltiples medidas, válidas y confiables.** Como todas las medidas e instrumentos actuales del aprendizaje y avance del alumno (pruebas estandarizadas, evaluaciones docentes, portafolios de actividades de los alumnos y observación, entre otros) pueden presentar fuentes potenciales de error y sesgo, debe elaborarse gradualmente un enfoque complementario que utilice evidencias válidas de múltiples fuentes para complementar los instrumentos actuales en México, estimar costos y determinar la creación de las capacidades y los instrumentos que se requieren. Por ejemplo, si se cuenta con un contenido y estándares de desempeño claros de lo que se espera que los alumnos sepan y sepan hacer, es posible mejorar las mediciones que reflejan el aprendizaje y crecimiento esperados de los alumnos.
- **En lo posible, el uso de los datos de desempeño del alumno debe acompañarse de mediciones complementarias y confiables del aprendizaje del alumno, conforme sean elaboradas, puestas a prueba y validadas.** La relativa importancia de los datos de los alumnos y de las evaluaciones basadas en la escuela o por parte de maestros, puede definirse de nuevo conforme lo necesiten los objetivos de las políticas y las consecuencias de las evaluaciones. Australia, Alberta (Canadá) y Hong Kong-China son ejemplos de sistemas con mejor desempeño que pretenden combinar evaluaciones estandarizadas con evaluaciones basadas en la escuela (es decir, elaboradas y calificadas localmente pero moderadas externamente), proyectos y trabajos extensos de alumnos.
- **Los datos de desempeño del alumno, como los de la evaluación anual ENLACE en México, pueden jugar un papel importante en la rendición de cuentas y en las acciones para la mejora escolar.** Las acciones en curso emprendidas por la SEP y las autoridades educativas estatales hacia la presentación y uso de ENLACE demuestran el alto grado de aceptación social y el potencial de ENLACE. Los datos de desempeño de alumnos, agrupados por grupo, escuela, zona o estado, pueden emplearse en modelos estadísticos de análisis estáticos, de mejora o de avance, dependiendo de las prioridades de las políticas públicas y programas en México.
- **Debe establecerse un programa específico de desarrollo para la evaluación ENLACE que tome en cuenta aspectos de la demanda cognitiva, la alineación curricular y la coherencia.** Se deben tomar en cuenta las mejores prácticas internacionales y evidencias sobre los estándares de alumnos y sobre el proceso mismo de aprendizaje. El desarrollo de ENLACE debe fijar etapas y metas claras que aborden consideraciones técnicas (como la equiparación vertical), administrativas (tales como identificadores y vínculos únicos de alumnos, maestros y escuelas) y logísticas (mejor supervisión en las pruebas).¹⁷ Con un uso más extendido de ENLACE a futuro se debe pensar, por ejemplo, en la posibilidad de mejorar la supervisión y la seguridad en la administración de las pruebas. El programa también debe tener una visión de largo plazo que tome en cuenta los mejores ejemplos internacionales de contenidos y estándares de desempeño. Conforme los estándares de contenido y desempeño se establezcan en México, los datos de desempeño de alumnos pueden usarse, junto con modelos analíticos (modelos de valor agregado, por ejemplo), para objetivos de políticas y programas específicos. A lo largo del proceso debe prestarse atención a la alineación y coherencia entre los estándares, la evaluación y el desarrollo profesional para los maestros. Una visión clara del marco de

evaluación en México debe permitir que haya propósitos distintos y, al mismo tiempo, complementarios de las diferentes evaluaciones (es decir, ENLACE, EXCALE o posibles evaluaciones basadas en la escuela) y considerar la manera como deban seguir desarrollándose en el futuro dentro de un marco nacional común.

- ***Con datos de desempeño del alumno y modelos adecuados de crecimiento, es posible identificar a quienes tienen bajo desempeño, alto desempeño y casos que ameriten observación adicional y mayor seguimiento.*** Conforme el proceso de evaluación y medición se establezca, las consecuencias, tales como los incentivos, una mayor observación y asistencia para escuelas y maestros, se asociarían con los resultados. Esto implica no sólo un desarrollo gradual del proceso, sino también la posibilidad de contar con consecuencias y respuestas a los resultados en múltiples etapas.

NOTAS

1. Como referencia, el desempeño promedio de los alumnos en Alberta en PISA de 2006 fue significativamente superior al promedio canadiense, que ya se ubicaba entre los de mejor desempeño, junto con Hong Kong-China (OCDE, 2007; Bussiere, Knighton y Pennock, 2007). De un total de 57, Australia se ubicó entre las 10 economías con mejor desempeño (OCDE, 2007).
2. Las materias más comunes en las pruebas son matemáticas y el idioma nacional, ciencias sólo se incluye en siete de los 29 países en donde había información disponible (OCDE, 2008).
3. Con preguntas de respuesta conceptual, una vez que se formula la pregunta la carga de trabajo recae en los hombros del calificador; la correspondencia entre la puntuación asignada a un alumno particular y su estatus cognitivo depende de la destreza del calificador para detectar correctamente las señales que muestren ese estado cognitivo. Con preguntas de elección múltiple se ha realizado ya una cantidad considerable del trabajo para la evaluación, pues se ha dividido la actividad cognitiva en los pasos pertinentes con que deben identificarse los diversos niveles cognitivos mediante las distintas combinaciones de las opciones.
4. Para el gobierno de Alberta, Canadá, una de las responsabilidades profesionales de los maestros es la de calificar las pruebas de logro de los alumnos de la provincia (<http://education.alberta.ca/department/ipr/commission/report/reality/governance/bargmodel.aspx>), mientras que en un documento que presentó el mayor sindicato en el Reino Unido, el NASUWT, se menciona que la calificación de evaluaciones de los alumnos “no necesitan hacerla los maestros” (citado en Stevenson, 2004, p. 233).
5. Por ejemplo, con base en el Índice de Desarrollo ICT (*ICT Development Index*), que combina datos de acceso, uso y habilidades de la Unión Internacional de Telecomunicaciones (*International Telecommunication Union*), Corea del Sur se ubicó en el segundo lugar general, Finlandia en el noveno y México en el 75, por debajo de Chile (48), Turquía (59) y Brasil (60) (ITU, 2009).
6. Una fuente general sobre las pruebas proviene de los Estándares para la Examinación Educativa y Psicológica (*Standards for Educational and Psychological Testing*) (American Educational Research Association, American Psychological Association y National Council on Measurement in Education, 1999).
7. Esta sección se basa en gran medida en contribuciones de expertos que encargó la OCDE como parte del Acuerdo de Cooperación con el gobierno de México. Los dos documentos de trabajo son *Challenges and Opportunities for the Further Development of the ENLACE Assessment for Evaluation and Teacher Incentives in Mexico* (Zúñiga Molina y Gaviria, 2010), y *State-Level Teacher Evaluation and Incentive Practices in Mexico: Diagnostic Study* (Salieri et al., 2010).
8. Administrados por el Instituto Nacional para la Evaluación de la Educación (INEE), se trata de los Exámenes de la Calidad y el Logro Educativos (EXCALE). Los exámenes EXCALE son evaluaciones de alumnos, basadas en muestras administradas en ciclos de cuatro años, en ciertos grados claves en los niveles de preescolar, primaria y secundaria. La evaluación de 2011 se hará a los alumnos de tercer año de preescolar en español y matemáticas, seguida de un examen de español, matemáticas, ciencias naturales y ciencias sociales para alumnos de tercer año de secundaria en 2012.
9. Las escuelas se clasifican de acuerdo con los distintos programas de educación básica en el país. Con las clasificaciones se comparan los resultados de escuelas con características semejantes en términos de población estudiantil y de los recursos disponibles para alumnos y maestros.
10. La aplicación de la prueba ENLACE a la muestra control con propósitos de equiparación va acompañada de un cuestionario de contexto. La información relacionada con las características de la escuela se toma de las variables de estratificación empleadas para el proceso de muestreo. Si bien se han efectuado algunos estudios que relacionan estas variables contextuales con el desempeño en la evaluación ENLACE, no se ha hecho ningún otro uso operativo de esta información.
11. En esta evaluación no se incluyen consideraciones sobre la validez consecuente debido a los múltiples usos y amplitud de los propósitos a los que actualmente está sujeta la prueba ENLACE.
12. El estudio original es en español y se incluye como anexo en el documento que prepararon Zúñiga Molina y Gaviria (2010) para la OCDE, en el que se basó esta sección.
13. Hasta 2009, las pruebas de ENLACE de tercer año se diseñaron para reflejar el contenido acumulado del nivel de secundaria en su conjunto (es decir, las pruebas incluían también contenido del primero y segundo años). Esto puede explicar los hallazgos de Lizasoain Hernández y Joaristi Olariaga (2009).
14. Los valores de correlación entre las variables latentes de la evaluación ENLACE y SEP-ISA se encuentran en Zúñiga Molina y Gaviria (2010).
15. La *pertinencia* se refiere a la importancia relativa atribuida a cada tema por los expertos y que depende de la profundidad del tratamiento en los libros de texto de las distintas materias; la *credibilidad* se refiere a la viabilidad de elaborar preguntas de elección múltiple en relación con el contenido de las materias incluidas en cada prueba de aprendizaje particular; la *continuidad* se refiere a la medida en que el contenido específico forma parte de una secuencia de enseñanza que va más allá de un año-grado particular; la *amplitud* se refiere al nivel de inclusión de otro contenido asociado a niveles más bajos de complejidad (Zúñiga Molina y Gaviria, 2010).
16. Como la prueba ENLACE es anual, cada año se elabora una prueba paralela (llamada “forma previa a la prueba”) y se aplica a una muestra control de alumnos que también presentan la prueba normal dada a los alumnos ese año (esto se conoce como la “forma operativa” de ENLACE). Las preguntas de la prueba previa se ajustan junto con la forma operativa de la prueba y después se usan para elaborar la nueva prueba para el año siguiente.
17. Las recomendaciones técnicas, administrativas y logísticas específicas para el desarrollo de ENLACE y su posible papel en modelos de valor agregado se presentan en el capítulo 5.

Referencias bibliográficas

American Council on Education (Linn, R. L., ed., 1993), *Educational Measurement*, Oryx Press, Arizona.

American Educational Research Association, American Psychological Association y National Council on Measurement in Education (1999), *Standards for Educational and Psychological Testing*, American Educational Research Association, Washington, DC.

Baker, E. (2003), *Multiple Measures: Toward Tiered Systems*, University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Los Ángeles.

Baker, E. (2004), *Aligning Curriculum, Standards, and Assessments: Fulfilling the Promise of School Reform*, University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Los Ángeles.

Baker, E. (2010), "Assessment and Accountability", documento especializado encargado por la OCDE para el Acuerdo de Cooperación entre la OCDE y el gobierno de México.

Burstein, J.C. (2003), "The e-Rater Scoring Engine: Automated Essay Scoring with Natural Language Processing", en M.D. Shermis y J. Burstein (eds.), *Automated Essay Scoring: A Cross-Disciplinary Perspective*, Erlbaum, Mahwah, Nueva Jersey, pp. 113-122.

Chung, G.K.W.K. et al. (2001), "Knowledge Mapper Authoring System Prototype" (versión final entregada a la OERI), University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Los Ángeles.

Chung, G.K.W.K. y E.L. Baker (2003), "Issues in the Reliability and Validity of Automated Scoring of Constructed Responses", en M.D. Shermis y J. Burstein (eds.), *Automated Essay Scoring: A Cross-Disciplinary Perspective*, Lawrence Erlbaum, Mahwah, Nueva Jersey, pp. 23-40.

Darling-Hammond, L. y L. McCloskey (2008), "What Would it Mean to Be 'Internationally Competitive'?", Stanford University, consultado el 10 de mayo de 2010 en http://edpolicy.stanford.edu/pages/events/kerner/materials/intnl_assessment_pdk.pdf.

Lizasoain Hernández, L. y L. Joaristi Olariaga (2009), "Estudio de la dimensionalidad de las pruebas ENLACE (2008) mediante técnicas factoriales clásicas y métodos no paramétricos basados en TRI – Informe preliminar", incluido como anexo técnico V al documento especializado de Zúniga Molina y Gaviria (2010) que encargó la OCDE para el Acuerdo de Cooperación entre la OCDE y el gobierno de México.

Government of Alberta (Canada) – Education (2009), "The Alberta Student Assessment Study: Final Report", Edmonton, Alberta.

Instituto Nacional para la Evaluación de la Educación (INEE) (2010), *Panorama Educativo de México: Indicadores del Sistema Educativo Nacional 2009 Educación Básica*, INEE, Ciudad de México.

International Telecommunication Union (ITU) (2009), "Measuring the Information Society: The ICT Development Index", ITU, Ginebra.

Moriconi, G.M. (2009), "The Development Index of Basic Education and Teacher Evaluation in Brazil", presentación en el Taller Internacional OCDE/SEP "Towards a Teacher Evaluation Framework in Mexico: International Practices, Criteria, and Mechanisms", 1 y 2 de diciembre de 2009, Ciudad de México.

OCDE (2002), *PISA 2000 Technical Report*, OCDE, París.

OCDE (2005a), *PISA 2003 Technical Report*, OCDE, París.

OCDE (2005b), *Formative Assessment: Improving Learning in Secondary Classrooms*, OCDE, París.

OCDE (2007), *PISA 2006: Science Competencies for Tomorrow's World*, OCDE, París.

OCDE (2008), *Education at a Glance 2008: OECD Indicators*, OCDE, París.

OCDE (2009), *Assessment and Innovation in Education*, documento de trabajo de la OCDE núm. 24, OCDE, París.

OCDE (2010), *La medición del aprendizaje de los alumnos: Mejores prácticas para evaluar el valor agregado de las escuelas*, OCDE, París.

Parandekar, S.D., E. Amorim y A. Welsh (2008), “Prova Brasil – Building a Framework to Promote Learning Outcomes”, nota núm. 121, The World Bank, Washington, DC.

Salieri, G., L. Santibáñez y B. Naranjo (2010), *State-Level Teacher Evaluation and Incentive Practices in Mexico: Diagnostic Study*, estudio encargado por la OCDE para el Acuerdo de Cooperación entre la OCDE y el gobierno de México.

Stevenson, H. (2007), “Restructuring Teachers’ Work and Trade Union Responses in England: Bargaining for Change?”, *American Educational Research Journal*, vol. 44(2), pp. 224-251.

Zúñiga Molina, L. y J.L. Gaviria (2010), *Challenges and Opportunities for the Further Development of the ENLACE Assessment for Evaluation and Teacher Incentives in Mexico*, documento especializado encargado por la OCDE para el Acuerdo de Cooperación entre la OCDE y el gobierno de México.

CAPÍTULO 5

El valor agregado de las escuelas: Mediciones más justas y equitativas

5.1 Los modelos de valor agregado y la escuela como unidad de rendición de cuentas.....	70
5.2 Importancia de la calidad de los datos y la información.....	75
5.3 Consecuencias vinculadas a una evaluación más justa y fiable de escuelas y maestros.....	76
5.4 Consideraciones para México.....	76

El capítulo anterior presentó un resumen de los aspectos principales que giran en torno a las opciones de evaluación del alumno, y consideró los desafíos y oportunidades de la evaluación ENLACE en México. Sin embargo, para los sistemas educativos que efectúan evaluaciones estandarizadas periódicas a los alumnos, los datos de desempeño representan una oportunidad no sólo de tener una “instantánea” del desempeño de los alumnos con base en las puntuaciones brutas de las pruebas, sino también de explorar el crecimiento de los alumnos con el paso del tiempo. La experiencia muestra que los antecedentes de los alumnos tienen una gran influencia en el desempeño (McCall, Kingsbury y Olsen, 2004; OCDE, 2008). Las evaluaciones que no toman esto en cuenta corren el riesgo de hacer que quienes manifiesten un desempeño bajo, o alto, sólo reflejen las condiciones socioeconómicas de los alumnos y sus familias, así como sus características lingüísticas o étnicas.

Un método que emplea una cantidad limitada de países de la OCDE para evaluar de manera más justa y precisa las contribuciones de las escuelas a los resultados de aprendizaje y crecimiento de los alumnos, es el de los modelos de valor agregado (MVA). No sorprende que los sistemas con muchos años de experiencia en evaluaciones estandarizadas de los alumnos, como los de Reino Unido y los Estados Unidos de América, hayan implementado modelos de valor agregado. Para países que sólo hasta hace poco incorporaron pruebas de alumnos de gran escala, como México y Brasil, los MVA son una opción para definir objetivos de políticas públicas que tienden a mejorar los resultados educativos. Como se analiza en el capítulo 4, las inversiones de la Secretaría de Educación Pública (SEP) y las autoridades educativas estatales en ENLACE permitieron su amplia aceptación social (Zúñiga Molina y Gaviria, 2010; Salieri *et al.*, 2010). Dada la diversidad de los contextos educativos en México, las autoridades educativas federales y estatales pueden emplear métodos de valor agregado para fortalecer la rendición de cuentas y las acciones de mejora en la escuela. Sin embargo, el desarrollo de métodos de valor agregado requiere una consideración, un diseño y una planeación cuidadosos para que aborden de manera eficaz los desafíos respectivos.

Basado en el trabajo realizado por la OCDE sobre modelos de valor agregado y ejemplos de varios países (*i.e.* Noruega, Polonia, Eslovenia, Reino Unido y los Estados Unidos de América), este capítulo presenta un panorama de los métodos de valor agregado, y analiza las implicaciones de las políticas públicas, los elementos del diseño y los retos de implementación que deben tomarse en cuenta. El capítulo también destaca la importancia de la calidad de los datos y de la información para poder vincular las intervenciones y consecuencias educativas con las evaluaciones precisas de escuelas y maestros. Por último, concluye con consideraciones específicas para México sobre los métodos de valor agregado que le permitan incrementar la rendición de cuentas y mejorar el desempeño de las escuelas.

5.1 LOS MODELOS DE VALOR AGREGADO Y LA ESCUELA COMO UNIDAD DE RENDICIÓN DE CUENTAS

Los sistemas educativos que atienden a alumnos con antecedentes diversos enfrentan un reto más grande todavía al elaborar evaluaciones precisas y justas del desempeño escolar. Por ende, los sistemas educativos que operan en contextos de alta diversidad y desigualdad (*i.e.* la distribución del ingreso o la discriminación lingüística o étnica, por dar un ejemplo), deben implementar procedimientos de evaluación y medición que sean significativamente comparables en el ámbito nacional, y que al mismo tiempo tomen en cuenta estas disparidades. Por ejemplo, los resultados en ciencias en PISA de 2006 muestran que el antecedente socioeconómico tiene una importancia mayor para el logro del alumno en México que en el promedio de países (Base de datos de PISA de 2006). En estos contextos, los métodos de valor agregado ofrecen una vía alternativa para fortalecer la rendición de cuentas y mejorar el desempeño de las escuelas.¹

Como incluyen medidas previas del desempeño de los alumnos, los métodos de valor agregado contribuyen a abordar algunos desafíos y problemas comunes con otros métodos:

- Dependencia a las mediciones, tales como las puntuaciones promedio en exámenes estandarizados o el porcentaje de alumnos que avanzan a niveles educativos más altos, que no tomen en cuenta otros factores que influyen en el avance: habilidad innata de los alumnos, antecedentes socioeconómicos de los alumnos, influencia de compañeros e individuos dentro y fuera de las escuelas, sucesos o acontecimientos impactantes externos que afecten en gran medida el aprendizaje, y aleatoriedad general de las evaluaciones de los alumnos (OCDE, 2008).

- Considerar factores desapercibidos que contribuyen a la medición inicial del desempeño del alumno, tales como la capacidad o el coeficiente intelectual del alumno, que plantean dificultades para los modelos de logro contextualizado (OCDE, 2008; Raudenbush, 2004).
- Incentivos no deseados para que el personal de la escuela excluya a alumnos que se consideran de bajo desempeño, o para conservar sólo a los alumnos de mejor desempeño (OCDE, 2008; Wilson, 2004).

Los modelos de valor agregado abordan estos problemas porque incluyen medidas de desempeño del alumno conforme evolucionan, con el paso del tiempo, hacia los objetivos prescritos. Si bien las medidas de valor agregado se calculan para los alumnos en lo individual, así como para las áreas temáticas, los niveles de grado, las escuelas y otras entidades jurisdiccionales (*i.e.* zonas escolares, municipios o estados), el interés general se enfoca en la escuela como unidad básica para la rendición de cuentas.

Recuadro 5.1 Jurisdicciones en México

En México, la zona escolar es la designación administrativa aplicada a un grupo de escuelas con fines de supervisión y vigilancia administrativa. De igual modo, los municipios son una de las tres unidades básicas de gobierno que puede contribuir de manera significativa no sólo en la infraestructura y las condiciones materiales de las escuelas, sino también en los programas de reconocimiento social a escuelas y maestros, por citar un ejemplo.

Fuente: Salieri *et al.*, 2010.

Basadas en consultas a expertos y en una revisión de modelos de valor agregado y métodos relacionados en distintos países (OCDE, 2008, 2011), las definiciones clave que se emplean en este y otros reportes de la OCDE son las siguientes:

Valor agregado de las escuelas	La contribución de una escuela al crecimiento de los alumnos hacia objetivos educativos establecidos o prescritos (p. ej., el aprovechamiento cognitivo o niveles mínimos de alfabetización). Esta contribución está aislada de otros factores que contribuyen al progreso educativo de los alumnos.
Modelos de valor agregado	Una clase de modelos estadísticos que estiman las contribuciones de las escuelas al avance de los alumnos en objetivos educativos establecidos o prescritos (p. ej., el aprovechamiento cognitivo o niveles mínimos de alfabetización) medido en al menos dos momentos temporales.

Pese a los desafíos, un sistema educativo puede emprender pasos graduales hacia la implementación de un método de valor agregado. Estos pasos pueden considerar métodos estadísticos y analíticos que, aunque aborden algunos de los principales factores que se toman en cuenta en verdaderos métodos de valor agregado (*i.e.* antecedentes socioeconómicos de los alumnos), no deben considerarse modelos plenos de valor agregado. Los sistemas educativos pueden emplear los siguientes procedimientos y modelos como fases de transición en vista de establecer un marco evaluativo que use modelos de valor agregado:

- estratificación* de escuelas semejantes (basada en información socioeconómica y de otro tipo que sea pertinente) para comparar al interior de los grupos el promedio de puntuaciones brutas de los alumnos de la escuela;
- modelo de los efectos de la escuela* basado en los residuales utilizando un análisis multivariante de corte transversal, con regresiones o adiciones de puntuaciones promedio de las pruebas y características demográficas pertinentes de los alumnos;

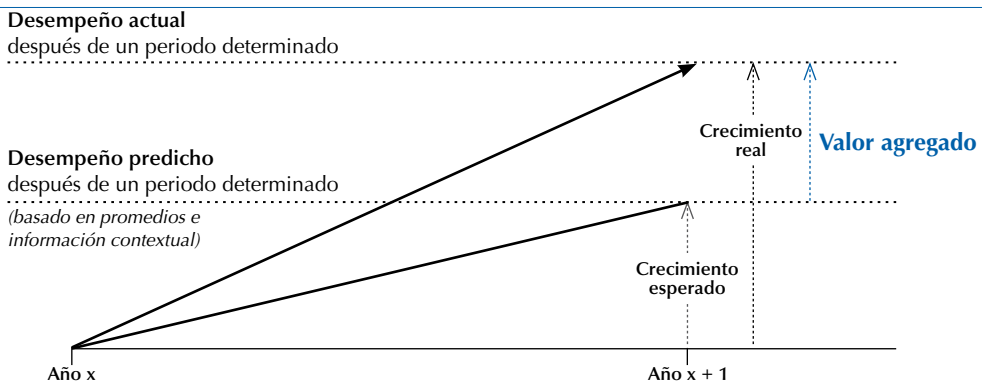
- c) *modelos de logro contextualizados* que estimen la magnitud de los factores que contribuyen al desempeño o logro del alumno utilizando una medición específica (es decir, en un momento determinado). Por lo común, el modelo efectúa una regresión de los antecedentes socioeconómicos de los alumnos, o de otras características contextuales, y la variable de un identificador de la escuela respecto a una medición de logro;
- d) *modelos de valor agregado* que usan al menos dos mediciones de desempeño del alumno hechas en dos momentos distintos; y
- e) *modelos de valor agregado contextualizado* que utilizan características contextuales pertinentes e importantes de los alumnos, tales como el género, el origen étnico y la escolaridad de los padres (OCDE, 2007). El diseño y uso de modelos que incluyan datos de los alumnos deben abordar el potencial de incentivos no deseados para la escuela o el personal administrativo que retenga la información.

Los sistemas no tienen que proceder necesariamente en este orden, ni aplicar todos los tipos de modelos. Algunos métodos pueden servir para un uso público del análisis; otros, con modelos más complejos, pueden ser usados de manera interna por parte de las autoridades educativas con fines de exploración u otros objetivos específicos, tales como la supervisión y evaluación de los programas. Los mismos modelos de valor agregado identifican no sólo a las escuelas de bajo desempeño, sino también a las de alto desempeño. Para los sistemas educativos que ofrezcan servicios en contextos muy variados y a poblaciones estudiantiles heterogéneas, los modelos de valor agregado ofrecen “una evaluación cuantitativa de la magnitud de la desventaja asociada a características particulares (*i.e.* el origen étnico, el ingreso, la escolaridad familiar, el idioma materno o la situación migratoria) en relación con el avance del alumno, y no sólo en relación con el logro del alumno en un momento determinado” (OCDE, 2008, p. 132).

Aunque aún no es una práctica común en países miembros y asociados de la OCDE, varios sistemas educativos emplean ya métodos de valor agregado para la rendición de cuentas y la mejora de escuelas. Dos ejemplos citados a menudo son el Sistema de Evaluación de Valor Agregado de Tennessee (*Tennessee Value-Added Assessment System*, TVAAS por sus siglas en inglés) en los Estados Unidos de América, y los Cuadros de Logro y Aprovechamiento (*Achievement and Attainment Tables*) en el Reino Unido. Otros estados y ciudades de los Estados Unidos de América implementaron modelos de valor agregado, así como países como Polonia, Noruega y Eslovenia. Además, hay al menos ocho países, incluyendo a México, con evaluaciones de alumnos que podrían ser usadas para explorar modelos de valor agregado, como se analizó en el capítulo 4.² De estos ocho países, sólo México y la Comunidad Flamenca de Bélgica producen datos anuales de evaluación de los alumnos (OCDE, 2008).

Gráfica 5.1

Representación esquemática de un modelo sencillo de valor agregado



Fuente: Adaptado de Martínez-Arias, Gaviria y Castro, 2009, y Goldschmidt *et al.*, 2005.

Las puntuaciones de valor agregado tienen una relación inherente con el desempeño de otras escuelas. Específicamente, la puntuación de una escuela en particular es un estimado de la diferencia entre la contribución de la escuela al aprendizaje de sus alumnos y la contribución *promedio* de todas las demás escuelas de donde se tomaron los datos usados en el modelo. Por ejemplo, el empleo de datos de otro grupo de escuelas daría por resultado distintas puntuaciones de valor agregado. La gráfica 5.1 proporciona una representación esquemática de un modelo sencillo de valor agregado. En la gráfica 5.1 se podría usar información contextual de los alumnos para ajustar los resultados predichos a partir del modelo ajustado (es decir, contribución promedio de la escuela al grupo de escuelas).

Se han elaborado y aplicado varios modelos de valor agregado: para distintos objetivos de políticas al interior de diferentes sistemas educativos, y para estimar efectos ya sean anuales o acumulados. Sin embargo, la mayoría se clasifica en cinco categorías generales: *i*) modelos de regresión lineal, *ii*) modelos jerárquicos de componente de varianza o efectos aleatorios, *iii*) modelos de efecto fijo, *iv*) modelos multivariados de efecto de respuesta aleatoria y *v*) modelos de curva de crecimiento latente. En el recuadro 5.2 hay una breve descripción de una forma común de estos modelos. En las etapas iniciales de la elaboración de modelos de valor agregado, las autoridades educativas y los órganos correspondientes deben explorar múltiples modelos y características de diseño para identificar el modelo de diseño más adecuado. Dicho modelo debe reflejar los objetivos claros de las políticas públicas, las muestras y la información contextual empleadas, así como las características de las evaluaciones de los alumnos.

Recuadro 5.2 Ejemplo de un modelo de valor agregado de regresión lineal

$$Y_{ij(2)} = a_0 + a_1 Y_{ij(1)} + b_1 X_{1ij} + \dots + b_p X_{pij} + \varepsilon_{ij}$$

i indica el subíndice para estudiantes dentro de las escuelas j ;

$Y_{ij(2)}$ = puntuación final de la prueba;

$Y_{ij(1)}$ = puntuación anterior de la prueba;

$\{X\}$ denota un conjunto de características de alumnos y familias;

$a_0, a_1, b_1, \dots, b_p$ denotan un conjunto de coeficientes de regresión; y

ε_{ij} denota desviaciones individuales y distribuidas normalmente con una varianza común para todos los alumnos.

En este modelo, si los alumnos de la escuela j alcanzan en la prueba puntuaciones finales más elevadas que el promedio, los residuos correspondientes tenderán a ser positivos, lo que generará un estimado de valor agregado positivo para la escuela. Para obtener estimados consistentes con este modelo, es necesario que los covariados incluidos *no se correlacionen* con el término de error, que puede incluir un efecto escolar además de los errores idiosincrásicos. Más aun, esto no toma en cuenta la estructura del término de error – como se hace en otros modelos.

Todos los indicadores empíricos del desempeño escolar están sujetos a la variabilidad y a sesgos. Como los métodos de valor agregado emplean fuentes empíricas distintas a las observaciones del salón de clase o a las visitas de inspección, ofrecen un valioso anclado cuantitativo de la evaluación del desempeño de una escuela. Al diseñarse métodos de valor agregado robustos, se deben abordar los siguientes aspectos estadísticos, metodológicos y de implementación:

- **Aspectos de diseño**

- Calidad de los datos de la evaluación y la prueba, incluyendo su alineación con metas curriculares (junto con las reformas pertinentes), así como la validez y confiabilidad de los instrumentos de evaluación.

- Integridad y cobertura de los datos brutos de la prueba y de los datos contextuales de los alumnos que se van a usar para los modelos contextualizados.
 - Validez y confiabilidad en la evaluación que resulta de los datos de los alumnos que van a usarse.³
 - Identificación clara de la variable dependiente que se va a usar en los modelos de valor agregado.
 - Tamaño de la escuela: en general, los resultados de escuelas con menos de 30 alumnos son más proclives a la inestabilidad, por lo que los resultados deben confirmarse mediante modelos exploratorios con datos e información específicos del país. Además, las escuelas con menos de 30 alumnos pueden agruparse para efectuar análisis de los agregados.
 - Complejidad técnica de los modelos, dada la disponibilidad de datos, y para la comunicación y comprensión del proceso por parte del personal escolar y el público en general.
 - Transparencia de los objetivos y los modelos, y comunicación de los resultados, para obtener la aceptación general. Esto necesita que las decisiones referentes a los resultados que las autoridades educativas van a usar de manera interna estén disponibles para las autoridades educativas locales, para las escuelas locales, que sean accesibles al público en general, o una combinación de las tres cosas.
 - Ponderación diferencial de los avances en las puntuaciones (*i.e.* en el extremo superior o en el inferior del desempeño).
 - Los costos de la recopilación y validación de los datos para crear bases de datos útiles suelen formar parte del presupuesto vigente de la evaluación que se emplea. Los costos de los modelos en sí y del análisis son, en cambio, relativamente modestos.
- **Aspectos estadísticos**
 - Ajustes para las características del alumno, de la escuela o contextuales, si el modelo los requiere (*i.e.* sin efectos fijos).
 - Varianza de los estimados de valor agregado, que también sirven para los intervalos de confianza de las puntuaciones de valor agregado de las escuelas.
 - La estabilidad intertemporal de puntuaciones de valor agregado también se ve afectada por los cambios de instrumento de evaluación que se emplee, los cambios de datos contextuales y la volatilidad de los resultados de escuelas más pequeñas con menos alumnos. En los casos en que haya una gran cantidad de escuelas pequeñas (es decir, con menos de 30 alumnos), se les puede agrupar por región geográfica o unidad administrativa para poder incluirlas en el análisis (OCDE, 2008, p. 196). En lo posible, se recomienda usar una media móvil de tres años en la puntuación de valor agregado por escuela para abordar problemas de inestabilidad.
 - El sesgo como medida de inexactitud y solidez fundamentales que sirva de punto de partida para suposiciones subyacentes respecto a la naturaleza de los datos empleados (*i.e.* omisión de ciertas variables), respecto a la estructura del modelo, o de ambas cosas. Esto reviste una particular importancia en sistemas educativos con una alta diversidad de alumnos pero que tienen una cantidad y una calidad limitadas de datos del nivel de los alumnos. Además, se debe tomar en cuenta si los índices de movilidad se correlacionan positivamente con alumnos de antecedentes u orígenes étnicos desfavorecidos.
 - Una media cuadrada de error, lo cual sugiere explorar modelos de efecto aleatorio; sin embargo, se consigue una menor variabilidad de estos modelos si se permite un sesgo mayor.
 - Los datos faltantes en el nivel de los alumnos, o en el de la escuela, deben ser abordados mediante evaluaciones iniciales de la calidad de los datos y las medidas correctivas correspondientes.

- El grado de semejanza entre los resultados de valor agregado producidos por diversos modelos. Al escoger el modelo que se va a emplear, los modelos exploratorios deben usar datos reales de los alumnos para identificar diferencias de puntuaciones producidas por distintos modelos y parámetros, así como sus implicaciones potenciales. Sin embargo, la experiencia de intentos de aplicación y estudios sugiere que deben favorecerse los modelos más sencillos, en particular en las etapas tempranas de su desarrollo, a menos de que haya ventajas claras de lo contrario, dependiendo del propósito de los modelos y las condiciones específicas del país (OCDE, 2011). También pueden usarse modelos más complejos de forma interna para vigilar el uso público de modelos más sencillos, o para identificar escuelas o jurisdicciones individuales que muestren diferencias significativas. Estas discrepancias pueden necesitar una mayor recopilación de datos para identificar su origen.
- **Aspectos de la implementación**
 - Establecer objetivos de políticas educativas y mediciones del desempeño de las escuelas.
 - Elegir un modelo adecuado de valor agregado o mixto.
 - Elaborar una base de datos adecuada.
 - Llevar a cabo un programa piloto riguroso y eficaz, tal vez con un subconjunto de escuelas: para determinar las características contextuales adecuadas que hay que incluir; los requerimientos de datos adicionales que se puedan necesitar; y para efectuar análisis de sensibilidad de los modelos seleccionados. Los aspectos de compromiso de los actores implicados, comunicación y creación de capacidad también deben abordarse durante la fase piloto.⁴
 - Supervisión continua de las puntuaciones de valor agregado generadas.
 - Costos relativos a la comunicación, información y creación de capacidad, junto con los actores implicados, respecto a los modelos y resultados de valor agregado.

5.2 IMPORTANCIA DE LA CALIDAD DE LOS DATOS Y LA INFORMACIÓN

Conforme los sistemas educativos invierten en mejorar la calidad y equidad de los servicios educativos, la eficiencia de sus inversiones se vuelve cada vez más importante. Esto ha dado un impulso claro en varios países para mejorar la calidad y cobertura de los sistemas de datos y de información educativos. Los sistemas de datos e información sólidos no sólo son un requisito, también son un proceso continuo de mejora para los modelos de valor agregado. Como en otras formas de análisis, una calidad deficiente de los datos genera una mayor varianza y sesgo de los resultados de valor agregado, lo que a su vez socava la credibilidad de las acciones y consecuencias vinculadas a los resultados.

Por lo tanto, para poder desarrollar modelos de valor agregado, se deben abordar la totalidad y la calidad de cuatro tipos principales de datos y de información:

- *Datos de evaluación de los alumnos*, con referencias cruzadas mediante identificadores del estudiante. Esto incluiría medidas compuestas de puntuaciones, si es el caso, así como medidas específicas (es decir, requerimientos mínimos de lectoescritura) y objetivos de desempeño.
- *Información contextual al nivel del alumno*, en particular si hay necesidad de analizar el desempeño de grupos específicos de escuelas o alumnos (*i.e.* de grupos étnicos o con distintos idiomas maternos). También debe abordarse el problema de la movilidad de los alumnos.⁵
- *Información sobre el ámbito escolar*, en particular si hay varios tipos y tamaños de escuelas. También se puede incluir información sobre el programa y las políticas públicas, en especial si se les va a analizar, supervisar y evaluar. Para sistemas con una alta diversidad de contextos escolares, esto ofrece la oportunidad de garantizar que los modelos de valor agregado tomen en cuenta comparaciones entre contextos escolares semejantes.
- *Información y reportes sobre la evaluación escolar* que ofrezcan posibles explicaciones de las puntuaciones de valor agregado, y que juntos brinden un panorama más completo del desempeño escolar al que se puedan vincular las acciones y las consecuencias.

5.3 CONSECUENCIAS VINCULADAS A UNA EVALUACIÓN MÁS JUSTA Y FIABLE DE ESCUELAS Y MAESTROS

Dentro de un marco de rendición de cuentas, las evaluaciones justas y fiables del desempeño escolar suelen tener por resultado acciones y consecuencias para los maestros. De igual modo, para las acciones de mejora de las escuelas, las evaluaciones deben no sólo ofrecer información al personal de la escuela sobre lo que funciona y cómo mejorarlo, sino también darles las oportunidades para hacerlo. Como proporcionan una medida precisa y equitativa del avance de los alumnos con el paso del tiempo, los modelos de valor agregado son útiles en ambos casos para evaluar el desempeño de las escuelas.

Las fases iniciales del establecimiento de un marco de rendición de cuentas que incluya modelos de valor agregado deben identificar oportunidades para las acciones de mejora de las escuelas. Es posible combinar los incentivos positivos que refuerzan y mejoran el desempeño de las escuelas, del personal y de los maestros, con más evaluaciones, asistencia y recursos para escuelas con bajo desempeño. Por ejemplo, el capítulo 7 presenta un análisis de los diversos tipos de incentivos con que se puede motivar a los maestros al usar la escuela como unidad de rendición de cuentas.

Conforme avance el sistema de rendición de cuentas y se establezcan las capacidades locales para ofrecer una asistencia adecuada, apoyo y opciones de desarrollo profesional a los maestros, se pueden incorporar poco a poco otras medidas vinculadas a los resultados de desempeño de las escuelas. En la elaboración de los modelos de valor agregado para apoyar la rendición de cuentas y las acciones de mejora de las escuelas, son importantes los siguientes puntos:

- Para que sean eficaces los incentivos y estímulos, es vital garantizar la credibilidad y equidad de las medidas, instrumentos y procedimientos del desempeño.
- Una mayor rendición de cuentas e iniciativas de mejora pueden ser implementadas de manera gradual dentro del mismo marco general de rendición de cuentas. Distintos tipos de incentivos y consecuencias se podrían vincular a los resultados de la evaluación.
- Los resultados de valor agregado pueden usarse junto con otras medidas y evidencias de desempeño escolar. La naturaleza complementaria de diversas fuentes de evidencias, incluyendo los resultados de valor agregado, se determinará por el desarrollo gradual de las medidas y capacidades. A la larga, las medidas de distintas fuentes ofrecerán una imagen más completa y exacta del desempeño escolar, así como del personal y de los procesos de la escuela.

El cuadro 5.1 presenta un resumen de algunos de los principales beneficios e implicaciones de políticas de los métodos de valor agregado para la rendición de cuentas y la mejora escolares (OCDE, 2010).⁶

5.4 CONSIDERACIONES PARA MÉXICO

A continuación se enlistan las principales consideraciones y recomendaciones para elaborar métodos de valor agregado que evalúen el desempeño escolar. Deben considerarse a la luz de las conclusiones y recomendaciones en otros capítulos de este informe, en particular del capítulo 4.

- *Por la amplia diversidad de contextos educativos entre y al interior de los estados mexicanos, los modelos de valor agregado ofrecen una medida justa y más exacta del crecimiento del alumno y del desempeño de la escuela.* Las acciones en curso de la SEP y las autoridades educativas estatales respecto a la presentación y uso de los resultados de ENLACE, son un buen punto de partida y pueden aprovecharse con los resultados de valor agregado de las escuelas. Los desafíos de diseñar, planear e implementar un sistema de evaluación para mejorar la rendición de cuentas y las escuelas con modelos de valor agregado deben abordarse con rigor en todas las etapas, incluyendo la fase inicial de movilización, análisis y aplicación de conocimientos de las reformas educativas en México.

Cuadro 5.1

Beneficios e implicaciones de los métodos de valor agregado para la rendición de cuentas y la mejora escolares

Beneficios generales	Implicaciones de las políticas educativas y posibles consecuencias
Ofrecer medidas precisas de aspectos de alto y bajo desempeño de un sistema educativo.	<ul style="list-style-type: none"> – Identificar escuelas con crecimiento significativo en el aprendizaje de los alumnos y, de forma indirecta, el personal de la escuela que pueda ser elegido para recibir reconocimientos, recompensas, oportunidades de desarrollo profesional, y para participar en la identificación de prácticas eficaces. – Identificar a las escuelas y al personal escolar que demuestren constantemente ganancias significativas en el aprendizaje del alumno o a las escuelas con más éxito en mejorar el desempeño de alumnos en desventaja para que proporcionen información, prácticas e ideas para elaborar estándares de enseñanza, basados en prácticas demostradas, dentro de un sistema educativo específico como el de México. – Emplear los resultados de valor agregado sirve para evaluar la calidad de evaluaciones subjetivas de las prácticas escolares y docentes, lo que operaría como punto de referencia cuantitativo del marco de evaluación. Algunos resultados de valor agregado desencadenarían, por ejemplo, procedimientos específicos de evaluación escolar. – Canalizar programas y financiamiento compensatorios para los alumnos, escuelas y personal docente con menor desempeño. Las puntuaciones bajas de valor agregado desencadenarían más acciones de evaluación y observación en los mismos centros de trabajo para identificar las causas del bajo desempeño y hacerles frente. – Analizar los resultados de valor agregado de jurisdicciones específicas (estados, distritos, zonas y municipios), programas educativos o grupos de alumnos de un origen étnico específico, por ejemplo, para identificar áreas y acciones prioritarias. – Emplear los análisis de valor agregado para supervisar y evaluar programas públicos específicos, o como parte de su fase piloto.
Mejorar la identificación y el análisis de lo que genera los resultados (es decir, buenas prácticas).	
Incrementar la calidad, la equidad y la transparencia de los sistemas de rendición de cuentas y evaluaciones escolares, lo cual contribuiría a proporcionar políticas educativas para mejorar las escuelas.	
Fortalecer el desarrollo de sistemas de información que permitan a las escuelas analizar y evaluar su desempeño y fortalecer las acciones de evaluación escolares.	
Hacer más eficaz el financiamiento educativo al permitir que los recursos (humanos, materiales y financieros) vayan a donde más se necesitan.	
Contribuir a combatir las desigualdades socioeconómicas de una sociedad que pueden ocultarse en el ámbito escolar con mediciones indiscriminadas y menos precisas de desempeño.	

Fuente: Consejo Asesor de la OCDE sobre Políticas de Evaluación e Incentivos Docentes, 2010.

- *La validez y confiabilidad de ENLACE ofrecen a México una valiosa oportunidad* para explotar el potencial de los datos de desempeño de los alumnos que genera cada año. Los datos de los alumnos y los modelos adecuados son útiles para lograr una mejor rendición de cuentas de las escuelas y los maestros, así como para realizar acciones de mejora, intervenciones educativas y canalizar recursos adicionales a donde se detecte un bajo desempeño.
- *Dadas las condiciones actuales del sistema educativo en México, los modelos de valor agregado pueden basarse sobre todo en la escuela como unidad de rendición de cuentas*, aunque también las zonas escolares, los grupos de alumnos, los municipios y estados pueden servir para el análisis y las acciones. La equiparación vertical debe estar entre los primeros aspectos técnicos por revisar en el subsecuente desarrollo de ENLACE. Asimismo, deben evaluarse la calidad y la disponibilidad de información contextual que puede ser usada para los modelos contextualizados de valor agregado.

- *Las primeras fases de la elaboración de modelos de valor agregado en México pueden centrarse en ejercicios analíticos de gabinete, con datos reales de los alumnos, para identificar las debilidades y fortalezas de distintos modelos de valor agregado. Incluso antes de aplicar métodos de valor agregado a los datos de desempeño de los alumnos, se puede agrupar a las escuelas en contextos socioeconómicos y emplear modelos de logro contextualizados como posibles precursores de análisis completos de valor agregado. El proceso de establecer modelos de valor agregado tiene varias fases:*
 - i) Estratificación de escuelas semejantes (basada en el tipo y en la información socioeconómica o de otro tipo pertinente) para comparar al interior de los grupos los resultados promedio de puntuaciones brutas. También se deben identificar y abordar los aspectos referentes a la calidad e integridad de los datos de las pruebas y de la información contextual.*
 - ii) Ejercicios internos con modelos de valor agregado a cargo de las autoridades educativas para seleccionar modelos y abordar problemas técnicos con los datos. Se sugiere utilizar un **una media móvil de tres años** para los modelos.*
 - iii) Información pública, conocimiento y compromiso entre los actores implicados sobre los méritos, retos y oportunidades de los modelos de valor agregado, a los que se podría vincular para lanzar de nuevo la evaluación ENLACE con un programa claro para su fortalecimiento.*
 - iv) Atribuir consecuencias (leves al principio) para las escuelas de bajo desempeño (mayor exploración, observación y asistencia), así como para las de alto desempeño. Los mismos análisis de valor agregado servirían a la SEP y a las autoridades educativas estatales para identificar escuelas que puedan tener maestros y prácticas que valga la pena imitar y establecer como modelos.*
- *Las autoridades educativas pueden emplear análisis MVA para supervisar y efectuar ensayos de evaluación de políticas, programas y jurisdicciones específicos. La supervisión y la evaluación de programas públicos, como el Programa Escuelas de Calidad, se basarían en métodos de valor agregado. Además, se pueden evaluar escuelas que atiendan a grupos particulares de alumnos, de distintos orígenes étnicos por ejemplo, para identificar las escuelas más o menos eficaces para contribuir al avance de esos alumnos. Asimismo, se efectuarían análisis del desempeño escolar para detectar diferencias entre y al interior de municipios, zonas escolares y estados.*

NOTAS

1. Hay un análisis detallado de los MVA en el ámbito escolar en *Measuring Improvements in Learning Outcomes – Best Practices to Assess the Value-Added of Schools* (OCDE, 2008) y en la edición actualizada en castellano de 2011 y publicación hermana de este informe *La medición del aprendizaje de alumnos: Mejores prácticas para evaluar el valor agregado de las escuelas* (OCDE, 2011). Este capítulo recurre en gran medida a ambas publicaciones.
2. Los otros países son Bélgica (Comunidad Flamenca), República Checa, Dinamarca, Francia, Portugal, España y Suecia, que también contribuyeron al reporte de la OCDE sobre MVA en 2008.
3. Esto abarca la representación adecuada de conceptos y la varianza limitada independiente de los conceptos. Esto puede ser particularmente importante cuando se vinculen de manera vertical las escalas de puntuación de pruebas para diferentes años, o cuando no se hayan abordado adecuadamente el potencial de comportamientos no deseados.
4. Esto puede incluir el uso de intervalos de confianza al presentar las puntuaciones de valor agregado contextualizado, aunque también requiere una educación y formación suficientes para mejorar su comprensión y uso por parte de los actores implicados en la educación y del público en general.
5. Un aspecto importante sobre el uso de datos del alumno es la existencia de leyes de privacidad que pueden limitar el empleo de esta información para modelos de valor agregado, como en Polonia, o que puedan requerir el consentimiento firmado de los padres, como en Eslovenia (OCDE, 2008).
6. Un análisis detallado de los beneficios, características y aspectos de diseño se encuentra en la publicación hermana de la OCDE de 2011, actualizada y disponible en castellano, *La medición del aprendizaje de los alumnos: Mejores prácticas para evaluar el valor agregado de las escuelas*.

Referencias bibliográficas

Goldschmidt, P., P. Roschewski, K. Choi, W. Auty, S. Hebbler, R. Blank y A. Williams (2005), "Policymakers' Guide to Growth Models for School Accountability: How do Accountability Models Differ?", documento encargado por el Councils of Chief State School Officers, Washington, DC.

Martínez-Arias, R., J.L. Gaviria y M. Castro (2009), "Concepto y evolución de los modelos de valor añadido en educación", en *Revista de Educación*, vol. 348, pp. 15-34.

McCall, M.S., G.G. Kingsbury y A. Olson (2004), *Individual Growth and School Success*, Northwest Evaluation Association, Lake Oswego, Oregon.

OCDE (2007), *Learning for Tomorrow*, OCDE, París.

OCDE (2008), *Measuring Improvements in Learning Outcomes – Best Practices to Assess the Value-Added of Schools*, OCDE, París.

OCDE (2011), *La medición del aprendizaje de los alumnos: Mejores prácticas para evaluar el valor agregado de las escuelas*, OCDE, París.

Raudenbush, S. W. (2004), "Schooling, Statistics, and Poverty: Can We Measure School Improvement?", Educational Testing Service, Princeton, Nueva Jersey.

Salieri, G., L.G. Santibáñez y B. Naranjo (2010), *State-Level Teacher Evaluation and Incentive Practices in Mexico: Diagnostic Study*, estudio encargado por la OCDE para el Acuerdo de Cooperación entre la OCDE y el gobierno de México.

Wilson, D. (2004), "Which Ranking? The Impact of a 'Value-Added' Measure of Secondary School Performance", en *Public Money and Management*, enero, pp. 37-45.

Zúñiga Molina, L. y J.L. Gaviria (2010), *Challenges and Opportunities for the Further Development of the ENLACE Assessment for Evaluation and Teacher Incentives in Mexico*, documento especializado encargado por la OCDE para el Acuerdo de Cooperación entre la OCDE y el gobierno de México.

CAPÍTULO 6

La evaluación de docentes en servicio: Aspectos de las políticas y de la implementación

6.1 Prácticas internacionales	82
6.2 Cuatro preguntas clave que deben abordar los sistemas de evaluación	84
6.3 Consideraciones para México	85

Como se vio en los capítulos previos, los sistemas educativos deben permitir que todos los niños tengan acceso a la educación básica y mejorar el aprendizaje del alumno. Este capítulo aborda un factor clave para mejorar el aprendizaje del alumno: la calidad de la enseñanza. Las investigaciones muestran con claridad que la calidad de la enseñanza, y por ende el desempeño individual de cada maestro, es el factor que más influye el avance del alumno (Manzi y Sclafani, 2010; OCDE, 2009b). Al mismo tiempo, hay evidencias convincentes, como se vio en el capítulo 2, de que un mayor logro educativo se relaciona de manera estrecha con el crecimiento económico, beneficiando así tanto a la sociedad como al individuo (OCDE, 2010a). Por tal motivo, los sistemas de evaluación docente deben contribuir a garantizar que todo salón de clases cuente con un maestro eficaz, incluso en los ambientes más adversos.¹ El capítulo comienza con una breve revisión de algunos de los principales elementos de los sistemas de evaluación docente basados en las prácticas internacionales. Después considera las dimensiones y aspectos básicos de políticas comunes en la implementación de sistemas de evaluación docente. Por último, el capítulo concluye con una serie de consideraciones y recomendaciones para que México apoye las acciones actuales y futuras con el fin de establecer un sistema eficaz de evaluación de maestros en servicio.

6.1 PRÁCTICAS INTERNACIONALES

Crear una planta laboral de educadores profesionales bien calificada es central para que un país sea capaz de mejorar los resultados escolares de sus jóvenes (Manzi y Sclafani, 2010; OCDE, 2005). La mejora y rendición de cuentas continuas requieren sistemas sólidos y precisos de datos y mediciones que permitan no sólo el registro del avance de alumnos y escuelas, sino también intervenciones oportunas con el apoyo adecuado. Debe destacarse que la rendición pública de cuentas implica igualmente la responsabilidad de las entidades educativas fuera de la escuela (*i.e.* distritos, gobiernos estatales). Las escuelas, en particular en ambientes muy difíciles, tal vez no estén en posibilidades de resolver por sí mismas las fallas y los problemas relativos a aspectos sistémicos o contextuales. En muchos países, incluso en los más ricos del mundo, como los Estados Unidos de América, existen desigualdades entre las escuelas en términos no sólo de los antecedentes y necesidades de los alumnos, sino también en los recursos y calificaciones profesionales del personal escolar que pueden atraer y conservar (OCDE, 2009b; NAE, 2009).

Así, la evaluación docente en todos los países de la OCDE forma parte de un marco más amplio de rendición de cuentas respecto a la eficacia de los sistemas, instituciones y actores educativos (OCDE, 2007). Dentro de un contexto mayor de rendición pública de cuentas, las evaluaciones docentes justas y eficaces brindan información crucial para la mejora y apoyos adicionales. Como todo el sistema y todos los actores necesitan rendir cuentas del aprendizaje y el crecimiento de los alumnos, las iniciativas de evaluación, incluyendo la evaluación docente, deben ser parte de este amplio mecanismo de acciones, recursos y objetivos alineados. En este contexto, la evaluación docente también opera como mecanismo de garantía de calidad al ofrecer una imagen de diagnóstico de los niveles actuales de desempeño, y al proporcionar evidencias para la toma de decisiones (OCDE, 2007).

La tendencia reciente de las reformas en muchos países apunta a la rendición de cuentas basada en las pruebas y en el desempeño (Sahlberg, 2009) como forma de asegurar que las reformas generales sean justas y proporcionen evidencias que las apoyen. Sin embargo, para que la evaluación sea eficaz, es importante que las políticas se basen en una responsabilidad y una confianza compartidas (Sahlberg, 2009).² Esto es en particular pertinente para México, en donde un sistema robusto de evaluación de maestros en servicio que emplea una gran variedad de instrumentos para medir el desempeño docente, también podría fomentar una cultura que valore la profesión docente. En este tipo de sistema, cada escuela y cada maestro sigue buenas prácticas de enseñanza, cumple con las expectativas planteadas y así, la mejora continua es una actividad cotidiana.

Con el fin de que los maestros sepan en qué áreas deben centrarse para mejorar, y en qué consiste una “buena” práctica de enseñanza, la evaluación sumativa basada en expectativas y estándares de enseñanza claros aporta información importante. Los resultados de las evaluaciones sumativas pueden servir como fuente importante de evidencias para que los maestros rindan cuentas de las expectativas y el desempeño profesional (OCDE, 2007). Una concepción clara de lo que se considera “buena enseñanza”, y la creación de estándares de enseñanza, son fundamentales para la elaboración de un sistema de evaluación docente.

Las acciones de reformas, recientes y actuales, que se llevan a cabo en los Estados Unidos de América ofrecen un ejemplo ilustrativo de los retos y problemas que enfrentan los distintos niveles de gobierno y las escuelas en su intento de mejorar el logro y la rendición de cuentas, y de proporcionar también resultados justos y precisos de la evaluación docente. Aunque los Estados Unidos de América no han tenido el mejor desempeño en PISA, sí brindan un ejemplo valioso sobre la forma en que las jurisdicciones locales, los estados y los distritos escolares pueden elaborar sistemas de evaluación docente conforme a directrices más amplias y del ámbito federal. Por ejemplo, el caso del estado de Delaware presenta un enfoque interesante del uso de los datos de los resultados del aprendizaje del alumno a través de su Sistema de Evaluación del Desempeño (*Performance Appraisal System*), en el que no se puede clasificar a los maestros como eficientes o mejores a menos de que sus alumnos demuestren niveles satisfactorios de crecimiento del aprendizaje (*Delaware Department of Education*, 2010).

Otro aspecto importante respecto a los estándares y la evaluación es la coherencia. Específicamente, es esencial la coherencia entre los estándares curriculares y del desempeño, los estándares de la buena enseñanza, así como de la evaluación y el desarrollo profesional (NAE, 2009). En este sentido, la elaboración de estándares de enseñanza es un paso importante para un enfoque basado en estándares que pretenda mejorar el desempeño del sistema educativo mediante la rendición de cuentas.³ El Reino Unido ofrece un buen ejemplo de coherencia en la alineación de la evaluación de resultados del alumno y las prácticas de enseñanza referentes al currículo (*Qualifications and Curriculum Development Agency*, 2010).

El diseño de los esquemas de evaluación docente plantea retos a todos los países con una variedad de problemas que necesitan una solución (Manzi y Sclafani, 2010):

- ¿Cuáles deben ser los componentes de un sistema justo de evaluación docente?
- ¿Cómo deben equilibrarse los propósitos formativos y sumativos de la evaluación?
- ¿Cómo deben comprometerse los maestros en el diseño y la implementación de los sistemas de evaluación docente?
- ¿Cómo elaborar, implementar y evaluar estándares confiables de la práctica de la enseñanza para que sirvan de base a las mismas evaluaciones?
- ¿Cómo usar los resultados de la evaluación del alumno para evaluar a los maestros?
- ¿Qué clase de compensaciones o consecuencias deben asociarse a los resultados de las evaluaciones docentes?

Los países cuentan con distintos enfoques para evaluar a los maestros en servicio (OCDE, 2007).⁴ Algunos lo hacen sobre todo con fines formativos,⁵ y se centran tanto en la identificación de debilidades en la práctica de la enseñanza de maestros en lo individual, como en el apoyo a las mejoras. Otros países los evalúan con propósitos sumativos al asociar algunas consecuencias para los maestros de acuerdo con los resultados de la evaluación; algunas más con ambos fines.

Sin embargo, lo que queda claro es que la evaluación de maestros en servicio mejora el desempeño docente (Barber y Mourshed, 2007). Por ejemplo, en Chile la experiencia demuestra una relación positiva entre los promedios de los alumnos en la prueba de logro del alumno SIMCE (Sistema de Medición de Calidad de la Educación) y la cantidad de maestros con buen desempeño en las evaluaciones docentes (Manzi y Sclafani, 2010). La evaluación de los maestros mejora la práctica docente porque identifica a los maestros eficientes y a los que necesitan apoyo. También contribuye a diseñar mejor los programas de desarrollo docente y a retener a los buenos maestros, proporcionando así herramientas para el diseño y provisión de incentivos y esquemas de pago (OCDE, 2009b). La coherencia entre un sistema de evaluación docente y uno de opciones continuas para generar capacidades y recompensas es, en este sentido, un equilibrio complejo pero crucial que los países se esfuerzan por alcanzar.

Y sí, la tarea es compleja. Además de una coherencia y una planeación cabal, la evaluación docente requiere una implementación reflexiva y cuidadosa, si no, es poco probable que tenga un gran impacto en el desempeño de los alumnos. Como se analizó en el capítulo 2, la implementación eficaz puede ser difícil. La siguiente sección presenta áreas de políticas públicas y problemas comunes que deben atenderse.

6.2 CUATRO PREGUNTAS CLAVE QUE DEBEN ABORDAR LOS SISTEMAS DE EVALUACIÓN

Deben considerarse cuatro áreas al diseñar un sistema de evaluación docente (Mancera y Schmelkes, 2010), las cuales se presentan a continuación:

¿Por qué evaluar?

Los dos principales objetivos de la evaluación docente son obtener buenos resultados educativos, que es la meta final de la enseñanza, y evaluar el proceso de enseñanza. La evaluación docente pretende asegurar que los maestros se desempeñen a toda su capacidad para mejorar el aprendizaje de los alumnos. Al mismo tiempo, busca mejorar la propia práctica del maestro al identificar fortalezas y debilidades para conseguir un mayor desarrollo profesional. Estos dos enfoques suelen referirse respectivamente a evaluaciones sumativas y formativas (OCDE, 2007). Los resultados educativos dependen de muchos factores, pero hay tres criterios principales, que son: contar con maestros eficientes, escuelas eficaces y un liderazgo escolar eficaz. La reforma sistémica debe hacer que se vea a la escuela como la unidad para la rendición de cuentas.

¿Qué evaluar?

Los sistemas de evaluación docente deben ser capaces de identificar a los maestros eficientes y las prácticas eficaces de enseñanza. Como la meta final del sistema educativo es el aprendizaje del alumno, se deben tomar en cuenta sus resultados. Por ende, un sistema de evaluación docente necesita estándares de una buena enseñanza y un marco de evaluación cabal y bien planeado. Es vital la participación de todos los actores implicados, y en especial de los maestros, en el diseño del marco para que sea fructífero. La evaluación debe acompañarse de retroalimentación y apoyo para que todos los maestros tengan la capacidad de mejorar su desempeño. Más importante aún, debe quedar claro el nexo entre el sistema de evaluación y el desarrollo profesional. De acuerdo con Danielson (2007), se deben incluir las siguientes áreas en el marco de la evaluación docente: la planeación y la preparación; el ambiente en el salón de clases; la instrucción; y las responsabilidades profesionales.

La experiencia internacional también muestra que un sistema de evaluación docente debe partir de estándares sólidos de lo que es una buena enseñanza. Los estándares deben tener ciertas características (Mancera y Schmelkes, 2010): *i*) cubrir todos las áreas definidas de la enseñanza, *ii*) establecer niveles de competencia para cada aspecto específico que defina los dominios de la labor docente y escolar, *iii*) reflejar un grupo nodal de desempeño que debe ser observable en todos los maestros y todas las escuelas, *iv*) definir y hacer operacionales las metas y resultados deseados de una buena enseñanza, y *v*) ser dinámicos para permitir una revisión continua de los estándares, de modo que permanezcan escalados con precisión y tomen en cuenta todos los aspectos de las prácticas de enseñanza. Para garantizar el éxito de una evaluación, los maestros necesitan participar no sólo en la creación de los estándares, sino también en formarse eficazmente.

¿Cómo evaluar?

El reto es diseñar un sistema justo, transparente, objetivo y creíble para los maestros. Por ende es crucial crear un sistema de evaluación con diversos instrumentos que midan el desempeño docente, y mecanismos para cruzar referencias de la información que observen a los maestros desde varios ángulos, permitiendo así que se juzgue el desempeño de los maestros de la manera más objetiva posible al cubrir la mayoría de los aspectos de la profesión docente. Algunos instrumentos claves son el desempeño del alumno, portafolios del alumno, autoevaluación, entrevistas y pruebas de conocimientos. Las visitas escolares y la observación en el salón de clases contribuyen a triangular y a validar resultados con distintos instrumentos. Un análisis comparativo de las prácticas de evaluación docente de Manzi y Sclafani (2010) pone de relieve algunos aspectos importantes al respecto. Por ejemplo, ambos observaron que la práctica internacional difiere en los instrumentos empleados y que las observaciones en el salón de clases sirven para contrarrestar el hecho de que, en algunos países, las pruebas estandarizadas no cubran todas las áreas que enseñan los maestros. La experiencia sugiere asimismo que, si bien es más fácil para los directores reconocer a los maestros cuyos alumnos presentaron un logro

bajo o alto en pruebas estandarizadas, en cambio tienen una mayor dificultad en juzgar a los maestros que se encuentran en un lugar intermedio (OCDE, 2007).

¿Quién evalúa?

Debe garantizarse la disponibilidad de evaluadores capacitados y competentes. Los evaluadores eficientes por lo menos deben: *i)* conocer la labor que efectúan los maestros, *ii)* estar capacitados para hacer las observaciones esperadas y *iii)* ser autónomos respecto al maestro evaluado (Mancera y Schmelkes, 2010).

La equidad es también un factor importante que hay que considerar cuando se elabora un sistema de evaluación. Aunque el marco completo de un sistema de evaluación permite el establecimiento de medidas comunes de desempeño para todos los maestros, también debe reconocer las situaciones tan distintas en las que laboran. Esto es especialmente cierto en países con grandes disparidades, como es el caso de los Estados Unidos de América y de México. Otros aspectos que necesitan tomarse en cuenta son el estado de la profesión docente, las inquietudes sobre la calidad y justicia de la educación, los problemas de retención, la tensión entre evaluaciones internas y externas, la definición de la evaluación docente como tal⁶ y una rendición de cuentas relativamente débil del aprendizaje del alumno.

6.3 CONSIDERACIONES PARA MÉXICO

Las acciones de reforma en curso en México que se centran en los aspectos de la selección de maestros, la rendición de cuentas y las evaluaciones pueden ofrecer ideas útiles para países que enfrentan contextos y dificultades semejantes: una geografía heterogénea, desigualdades en el ingreso, descentralización y procesos de devolución continuos, y un sistema de educación básica (preescolar, primaria y secundaria) grande y complejo. En este contexto, y para apoyar las acciones actuales y futuras, se presenta a continuación un resumen de recomendaciones para implementar un sistema de evaluación de maestros en servicio en México que permita a los propios maestros mejorar, ser reconocidos y contribuir a los resultados educativos generales en todos los niveles del espectro de desempeño:⁷

- *Es vital establecer un consenso entre los actores implicados sobre la importancia de elaborar un marco de evaluación de los maestros en servicio completo, transparente y justo.*
- *La base para este marco es la elaboración de estándares de enseñanza que brinden a los maestros una guía clara de lo que se considera una buena práctica de enseñanza, así como oportunidades para su desarrollo y mejora profesionales.⁸*
- *Es esencial garantizar que todos los maestros alcancen niveles mínimos de desempeño y resultados profesionales.*

Como señala en su Programa Sectorial de Educación 2007-2012, el gobierno mexicano ve la creación de un sistema de evaluación de maestros en servicio con base en estándares como una de sus prioridades en el área de la educación. El Programa indica que la evaluación es una herramienta central para asegurar la calidad de la educación. De tal forma, la evaluación, como herramienta de comunicación y como base para diseñar políticas públicas, resulta vital para la rendición de cuentas (SEP, 2007).

México ha realizado algunos progresos en la evaluación docente durante los años recientes, entre los que se incluye la Carrera Magisterial, el Escalafón Docente⁹ y algunas acciones exploratorias, con el fin de elaborar estándares para los maestros, cuyo conjunto se debe revisar al momento de diseñar un sistema completo de evaluación docente. Sin embargo, las acciones actuales no están necesariamente articuladas o completas, ni se basan en una definición aceptada de estándares de desempeño para alumnos o maestros. La eficacia de acciones anteriores también es importante. Un estudio a cargo de la Rand Corporation reveló que la Carrera Magisterial tenía poco o ningún impacto para incrementar el logro del alumno (Santibáñez *et al.*, 2007). Tradicionalmente, la evaluación ha sido responsabilidad de los directores escolares y, en menor grado, de los supervisores escolares u otras autoridades educativas. El cuadro 6.1 resume las prácticas de evaluación vigentes en México.

Cuadro 6.1

Panorama general de las prácticas de evaluación docente en México

Función	Instrumento	Propósito
Acceso a la educación docente (Escuelas Normales)	Varios	Selección/Admisión
Certificación de estudios iniciales	Examen presentado al final de la formación docente inicial – EGEU/CENEVAL (Examen de término de los estudios en Escuelas Normales)	Formativo/Sumativo
Selección	Examen de selección (Concurso nacional de oposición para la obtención de plazas docentes)	Sumativo = Para obtener una plaza docente
Promoción horizontal (formación continua)	Examen – Carrera Magisterial/Examen de preparación profesional Examen – Carrera Magisterial/ENAMS (Exámenes Nacionales para la Actualización de los Maestros en Servicio) Examen (para calcular los resultados de aprendizaje de los alumnos) – Carrera Magisterial/Evaluación de Aprovechamiento Escolar	Sumativo Formativo/Sumativo Formativo/Sumativo
Reconocimiento y estímulos	Modelo de cálculo (Programa de Estímulos a la Calidad Docente) (basado en los resultados de ENLACE, los tipos de escuelas y el contexto socioeconómico)	De reconocimiento, con la intención de convertirse en programa de incentivos a futuro

Fuente: Traducido y adaptado de Zorrilla, 2009.

Cuadro 6.2

Implementación de pasos en orden secuencial

Pasos de implementación	
A. Pasos iniciales generales	<ol style="list-style-type: none"> 1. Presentar un sistema de evaluación docente 2. Hacer partícipes a los actores implicados 3. Hacer partícipes a las autoridades locales 4. Identificar a un defensor para el plan 5. Garantizar el financiamiento
B. Creación del marco de evaluación	<ol style="list-style-type: none"> 1. Elaborar estándares para la enseñanza 2. Elaborar medidas válidas de desempeño 3. Crear un sistema robusto de manejo de datos 4. Formar evaluadores competentes
C. Preparación en las escuelas	<ol style="list-style-type: none"> 1. Formar a líderes escolares y maestros para el sistema de evaluación 2. Crear acceso a la retroalimentación y mejora 3. Implementar un plan de comunicación para maestros y directores
D. Fase piloto	<ol style="list-style-type: none"> 1. Diseñar un programa piloto para poner a prueba el diseño, los instrumentos y los evaluadores
E. Implementación completa	<ol style="list-style-type: none"> 1. Preparar las transiciones de la fase piloto a la implementación completa 2. Vincular consecuencias al sistema de evaluación 3. Crear un plan de evaluación
F. Monitoreo y evaluación	<ol style="list-style-type: none"> 1. Definir indicadores 2. Establecer líneas base 3. Definir fuentes y procesos de información

Fuentes: Elaborado y adaptado de Mancera y Schmelkes, 2010, y OCDE, 2009b.

Por tanto, una cuestión clave es la forma en que México puede comenzar a crear de manera efectiva un sistema completo de evaluación de los maestros en servicio. Como se analizó en el capítulo 2, un paso fundamental es aprender de las experiencias internacionales – pero adaptándolas a las condiciones, limitaciones y oportunidades del sistema educativo mexicano (que también sugieren Mancera y Schmelkes, 2010). Para lograrlo, Mancera y Schmelkes (2010) recomiendan la implementación de etapas en esencia secuenciales, aunque también pueden echarse a andar al mismo tiempo, dependiendo de las circunstancias y las oportunidades (cuadro 6.2).

Las evidencias sobre el desempeño de los maestros y los resultados de aprendizaje de los alumnos, deben provenir de fuentes diversas. Además de entrevistas, portafolios, observaciones en el salón de clases y pruebas de conocimientos a los maestros, es fundamental el empleo de pruebas estandarizadas a los alumnos para evaluar el desempeño de los maestros. Como se analizó en los capítulos previos, México ya aplica una evaluación nacional a los alumnos, ENLACE, que puede servir para identificar escuelas con desempeño docente superior al promedio. ENLACE es un activo importante del sistema educativo mexicano y puede ser útil como parte de un proceso colectivo de evaluación en la que la escuela funge como unidad de rendición de cuentas. Para las evaluaciones sumativas individuales de los maestros, ENLACE necesitaría evolucionar para incluir componentes de valor agregado, de modo que mida con más claridad la contribución de las escuelas y, a la larga, la contribución de los maestros al aprendizaje de sus alumnos en contextos específicos.

Diseñar, poner a prueba e implementar un sistema completo, transparente y justo de evaluación de los maestros en servicio debe ser un proceso gradual. Chile se tardó diez años en diseñar e implementar su sistema, y los resultados educativos todavía siguen rezagados. México tiene también una clara necesidad de mejorar el desempeño de sus alumnos. Con base en análisis del Consejo Asesor de la OCDE sobre Políticas de Evaluación e Incentivos Docentes, y para que México avance al respecto, Mancera y Schmelkes (2010) hicieron 11 recomendaciones que se enlistan en el cuadro 6.3.

Cuadro 6.3

Resumen de recomendaciones específicas para México sobre la evaluación docente

Recomendación 1.	Establecer una estructura de liderazgo y reglas claras para la gobernanza del sistema de evaluación.
Recomendación 2.	Establecer una unidad técnica responsable de la implementación de la evaluación.
Recomendación 3.	Elaborar estándares de enseñanza.
Recomendación 4.	Diseñar un modelo de evaluación de maestros en servicio que evolucione poco a poco y pase de ser de un sistema sólo formativo a uno que combine aspectos tanto formativos como sumativos.
Recomendación 5.	Definir los instrumentos para el sistema de evaluación de los maestros en servicio.
Recomendación 6.	Elaborar un sistema de apoyo para el desarrollo basado en la escuela que genere la mejora de la práctica docente.
Recomendación 7.	Formar evaluadores.
Recomendación 8.	Reducir las cargas administrativas de supervisores y directores, y aumentar la autonomía escolar.
Recomendación 9.	Preparar un programa que permita a ENLACE medir el valor agregado.
Recomendación 10.	Obtener el respaldo de los actores implicados claves para establecer el sistema de evaluación docente.
Recomendación 11.	Poner a prueba y evaluar el diseño, los instrumentos y los evaluadores en distintos contextos, antes de echar a andar el sistema de evaluación en todo el sistema escolar.

Fuente: Mancera y Schmelkes, 2010.

Además de la creación de estándares, México debe llegar a un consenso sobre la importancia de diseñar e implementar un sistema completo, transparente y justo de evaluación de los maestros en servicio. Crear y operar un marco para la evaluación docente es una empresa prolongada y compleja. Es necesario sortear los cambios políticos y administrativos y hacer partícipes del proceso a los actores implicados. Las autoridades, tanto nacionales como estatales, deben ser parte activa del diseño del marco, junto con la participación de las autoridades locales. Los sindicatos magisteriales y la sociedad civil también deben participar en la toma de decisiones importantes. Todos los actores importantes deben estar representados en un órgano que sea designado responsable de garantizar la implementación de un sistema que reúna la confianza y el apoyo de los maestros y la sociedad en general. Es importante establecer mecanismos para la formación docente continua; en el caso de México, se debe establecer y poner a prueba una evaluación formativa antes de introducir consecuencias significativas para los maestros en lo individual (Mancera y Schmelkes, 2010).

En el contexto de aumentar la rendición de cuentas y ofrecer oportunidades para fomentar las capacidades y el desarrollo profesional de los maestros, es importante que todos ellos cumplan con niveles mínimos de desempeño y resultados profesionales. El crecimiento del aprendizaje del alumno debe ser uno de los criterios de evaluación. De esta manera, no se podrá considerar a un maestro como eficiente hasta que sus alumnos no demuestren niveles satisfactorios de crecimiento, mientras que a un maestro no se le podrá clasificar como ineficiente si sus alumnos muestran niveles satisfactorios de avance. También es esencial que se tomen en cuenta aspectos básicos como la asistencia, la puntualidad y el tiempo efectivo de clase, en las etapas tempranas del marco de evaluación docente como forma de hacer que todos los maestros se desempeñen con toda su capacidad. Incluir criterios básicos como éstos genera ganancias considerables y oportunas para el sistema de evaluación docente de manera rentable (es decir, asegurar que se cosechen primero todas las frutas que están “al alcance de la mano”).

Como se analizó en el capítulo 3, las evidencias de estudios en los Estados Unidos de América y en países en vías de desarrollo muestran que las ausencias de los maestros tienen un gran impacto en el aprendizaje del alumno, en especial en las áreas más desfavorecidas y las comunidades remotas. La Encuesta Internacional sobre Docencia y Aprendizaje (*Teaching and Learning International Survey*, TALIS por sus siglas en inglés) realizada por la OCDE señala que, para una cuarta parte de los maestros mexicanos, sólo entre un 40% y un 60% del tiempo de instrucción se dedica realmente a la enseñanza (OCDE, 2009a). Además, los directores reportan que el 70% de la enseñanza en su escuela se ve obstaculizada porque los maestros llegan tarde, no se presentan o no preparan bien sus clases (OCDE, 2009a). Teniendo en mente la disparidad socioeconómica de las condiciones que existen en todo México, el desempeño profesional básico, como base sólida sobre la cual construir capacidades, es uno de los principales problemas que deben resolver las políticas públicas y las evaluación docentes.

NOTAS

1. Este capítulo recurre a dos documentos especializados encargados como parte del Acuerdo de Cooperación entre México y la OCDE: “*Report on In-Service Teacher Evaluation and Development Practices in a Comparative Perspective*”, de Jorge Manzi y Susan Sclafani (2010), y “*Specific Policy Recommendations on the Development of a Comprehensive In-Service Teacher Evaluation Framework*”, de Carlos Mancera y Sylvia Schmelkes (2010).
2. Las reformas basadas en la responsabilidad y la confianza implican la creación gradual de una cultura de responsabilidad y confianza dentro del sistema educativo que valore el profesionalismo de maestros y directores al juzgar lo que es mejor para los alumnos y al reportar su progreso de aprendizaje. A menudo implican la canalización de recursos y el apoyo a escuelas y alumnos en riesgo de fracasar o de quedar rezagados.
3. Una referencia citada a menudo es *Framework for Teaching*, de C. Danielson, Perrenoud (2004), aunque otras también pueden ser pertinentes: *Rewards and Incentives Group* (2009), *Ontario Ministry of Education* (2009), Khim Ong (2008) y *Singapore Ministry of Education* (2006).
4. Los esquemas actuales de evaluación docente en los sistemas educativos de la OCDE adoptan múltiples formas, y pueden diferir en términos de: alcance y métodos de evaluación docente; criterios y estándares, de instrumentos de recopilación de datos de acuerdo con el contexto y tradición educativos, actores participantes en el diseño e implementación del sistema de evaluación, y con el propósito principal de la evaluación. También varían las consecuencias de las evaluaciones en las carreras de los maestros. Si bien el cuadro de promoción y la escala salarial únicos aún están muy extendidos, varios países vinculan su evaluación docente ya sea con el reconocimiento y las recompensas, financieras o no, o con oportunidades de desarrollo profesional (OCDE, 2009b).
5. Las evaluaciones formativas son esenciales para medir la calidad del desempeño docente y son un instrumento útil para apuntalar el desarrollo profesional de los maestros. Un sistema formativo de evaluación, diseñado y orientado sobre todo en mejorar la calidad de la instrucción, también sirve para identificar a los buenos maestros.
6. En Europa, una meta importante de las políticas públicas de enseñanza es atraer y retener a los mejores candidatos a la profesión. En América Latina, es mejorar la calidad de los maestros en función (Manzi y Sclafani, 2010).
7. Estas recomendaciones deben considerarse en el contexto de las recomendaciones de la OCDE sobre desarrollo profesional docente y liderazgo escolar (OCDE, 2010b), ya que la calidad de las escuelas tiene un impacto importante en la enseñanza. Hay una descripción más amplia de las políticas de enseñanza vigentes en México en el capítulo 3 de *Mejorar las escuelas: Estrategias para la acción en México* (OCDE, 2010b).
8. Los estándares de enseñanza también deben reflejar los estándares de contenido. Esto reviste una especial importancia en México debido a las acciones de reformas en curso en esta área.
9. Tanto la Carrera Magisterial como el Escalafón Docente sirven sobre todo como mecanismos relacionados con el ascenso. Véase Santibáñez *et al.* (2007) y OCDE (2010b).

Referencias bibliográficas

Alcázar, L., H. Rogers, N. Chaudhury, J. Hammer, M. Kremer y K. Muralidharan (2006), "Why are Teachers Absent? Probing Service Delivery in Peruvian Primary Schools?", en *International Journal of Educational Research*, vol. 45, pp. 117-136.

Barber, M. y M. Mourshed (2007), *How the World's Best-Performing School Systems Come Out on Top*, McKinsey & Company, Londres.

Clotfelter, C., H. Ladd y J. Vigdor (2008), "Are Teacher Absences Worth Worrying about in the U.S.?", National Center for Analysis of Longitudinal Data in Education Research, documento de trabajo 24 (versión de mayo de 2008), consultado en www.nber.org/papers/w13648.pdf.

Danielson, C. (2007), *Enhancing Professional Practice: A Framework for Teaching*, 2a. ed., Association for Supervision and Curriculum Development, Alexandria, Virginia.

Danielson, C. (2008), *The Handbook for Enhancing Professional Practice: Using the Framework for Teaching in Your School*, 2a. ed., Association for Supervision & Curriculum Development (ASCD), Alexandria, Virginia.

Delaware Department of Education (2010), "Race to the Top: Application for Funding", CFDA, núm. 84395A, Narrative, El estado de Delaware.

Khim Ong, K. et al. (2008), "Teacher Appraisal and its Outcomes in Singapore Primary Schools", en *Journal of Educational Administration*, pp. 39-54.

Mancera, C. y S. Schmelkes (2010), "Specific Policy Recommendations on the Development of a Comprehensive In-Service Teacher Evaluation Framework", OCDE, París.

Manzi, J. y S. Sclafani (2010), "Report on In-Service Teacher Evaluation and Development Practices in Comparative Perspective", OCDE, París.

Miller, R., R. Murnane y J. Willet (2007), *Do Teacher Absences Impact Student Achievement? Longitudinal Evidence from One Urban School District*, National Bureau of Economic Research, documento de trabajo 13356, consultado en www.nber.org/papers/w13356.pdf.

OCDE (2005), *Teachers Matter: Attracting, Developing and Retaining Effective Teachers*, OCDE, París.

OCDE (2007), "Teacher Evaluation: Current Practices in OECD Countries and Literature Review", M. Isoré, documento de trabajo sobre educación, OCDE, París.

OCDE (2009a), *Creating Effective Teaching and Learning Environments: First Results from TALIS*, OCDE, París.

OCDE (2009b), *Evaluating and Rewarding the Quality of Teachers*, OCDE, París.

OCDE (2010a), *The High Cost of Low Education Performance*, OCDE, París.

OCDE (2010b), *Mejorar las escuelas: Estrategias para la acción en México*, Directorado de Educación de la OCDE, Implementación de Políticas Públicas, OCDE, París, consultado en www.oecd.org/edu/calidadeducativa.

Ontario Ministry of Education (2009), "Overview of the Ontario Teacher Performance Appraisal (TPA) System".

Perrenoud, P. (2004), *Diez nuevas competencias para enseñar. Invitación al viaje*, Grao, Madrid.

Qualifications and Curriculum Development Agency (2010), *A Big Picture of the Secondary Curriculum*, Qualifications and Curriculum Development Agency, http://curriculum.qcda.gov.uk/uploads/BigPicture_sec_05_tcm8-15743.pdf.

Rewards and Incentives Group (2009), "Teachers' and Head Teacher's Performance Management: Guidance", consultado el 24 de abril de 2010 en www.teachernet.gov.uk/management/payandperformance/performancemanagement.

Rogers, F.H. y E. Vegas (2009), *No More Cutting Class? Reducing Teacher Absence and Providing Incentives for Performance*, World Bank Policy Research, documento de trabajo 4847, The World Bank, Washington, DC.

Sahlberg, P. (2009), *A Short History of Education Reform in Finland*, Helsinki, consultado el 24 de abril de 2010 en www.pasisahlberg.com/index.php?id=64.

Santibáñez, L., J. Martínez, A. Datar, P. McEwan, C. Setodji y R. Basuto-Dávila (2007), "Breaking Ground: Analysis of the Assessment System and Impact of Mexico's Teacher Incentive Program 'Carrera Magisterial'", reporte técnico RAND, RAND Corporation, Santa Mónica, California.

SEP (2007), *Programa Sectorial de Educación 2007-2012*, consultado el 24 de abril de 2010 en http://upepe.sep.gob.mx/prog_sec.pdf.

Singapore Ministry of Education (2006), "Singapore Staff Appraisal (Education Service)", en MOE, Singapore.

Zorrilla, M. (2009), presentado durante el *Taller OCDE-MÉXICO. Hacia un sistema de evaluación docente en México: Prácticas internacionales, criterios y mecanismos. Panorama general de prácticas de evaluación en México como parte de las políticas de profesionalización docente*, Ciudad de México, 1 y 2 de diciembre de 2009.

CAPÍTULO 7

Los incentivos para docentes en servicio

7.1 Tipos de incentivos docentes.....	95
7.2 Directrices nacionales e implementación local: Importancia del equilibrio adecuado	106
7.3 Programas piloto, monitoreo y evaluación de incentivos	107
7.4 Consideraciones para México	108
Apéndice 7A.....	111

Como se analizó brevemente en el capítulo anterior, las investigaciones confirman la importancia de la calidad de los maestros para la enseñanza de los alumnos (OCDE, 2009a, 2005). Una investigación en los Estados Unidos de América, realizada hace casi 20 años, reveló que los alumnos cuyos maestros se ubican en la parte superior del espectro de efectividad alcanzan hasta un año más de crecimiento en el aprendizaje, en comparación con los alumnos cuyo maestro se acerca al extremo inferior del mismo espectro (Hanushek, 1992). En otro estudio muy citado, basado en datos de Tennessee, Estados Unidos de América, los investigadores observaron que si asignaban a dos alumnos del segundo grado de educación primaria, con un desempeño semejante, a maestros de alto y bajo desempeño respectivamente por un período de tres años, la diferencia en su desempeño al final de este período podía alcanzar hasta los 54 puntos porcentuales (Sanders y Rivers, 1996). Las ideas más recientes sobre los sistemas educativos eficaces confirman que la calidad de un sistema educativo no puede exceder la calidad de sus maestros (OCDE, 2009a; McKinsey y Company, 2007). Sin embargo, en vista de la creciente competencia y la desaceleración económica internacionales, los gobiernos se ven obligados cada vez más a hacer más con menos. Además de atraer y retener a los profesionales más calificados en la profesión docente, un reto importante es la forma en que los sistemas educativos motivan y apoyan a los maestros en servicio para mejorar su desempeño y aumentar el logro de los alumnos.¹

Por ende, las políticas y programas que evalúan y recompensan a los maestros eficaces son cada vez más importantes en el marco de una reforma educativa. Esto sucede tanto en economías desarrolladas como en vías de desarrollo, en las que hay una mayor necesidad de usar eficiente y estratégicamente los sistemas de compensación como mecanismo para mejorar la calidad de la enseñanza y el aprendizaje de los alumnos. De hecho, se afirma que el sistema de compensación de una organización es su sistema más importante en cuanto al manejo de los recursos humanos (Ehrenberg y Milkovich, 1987; Lawler, 1981).

Si bien las recompensas por desempeño se han empleado de manera eficaz en otras áreas laborales, en particular en el sector privado (Lazear, 1996), su reciente aplicación en el sector educativo, sobre todo para los maestros, aún está en vías de exploración, supervisión y evaluación. Los responsables de formular políticas educativas consideran actualmente tanto los incentivos que no implican compensaciones, como aquellos que sí lo hagan.² Es por esto que cada vez más se toman en cuenta los incentivos para maestros en servicio, ya que reflejan la abrumadora importancia de los maestros, el interés en mejorar el logro de los alumnos mediante mejores prácticas de enseñanza y la rentabilidad de recompensar un desempeño superior.

Los incentivos tienen diversos propósitos específicos: motivar a los maestros a trabajar en escuelas con dificultades para atraer a más personal o para enseñar ciertas materias, a solicitar formación, o a asumir funciones y responsabilidades en la escuela (OCDE, 2009a). Para alentar las acciones de mejora, los sistemas educativos también pueden recompensar a escuelas y maestros por un desempeño aceptable o superior al promedio, mientras sancionan al mismo tiempo a quienes manifiesten un bajo desempeño. La naturaleza y el diseño exactos del esquema de incentivos docentes dependerán así de los objetivos establecidos por las políticas públicas. En este sentido, los incentivos son un elemento importante dentro de un marco más amplio de rendición de cuentas enfocado en el logro de los alumnos.

Este capítulo ofrece un panorama de las prácticas internacionales, previas y actuales, relacionadas con diferentes tipos de incentivos docentes, y pretende identificar las lecciones aprendidas y las mejores prácticas con el fin de elaborar políticas eficaces de incentivos docentes.³ Comienza así por detallar varios tipos de incentivos docentes, como los que se basan en insumos y los que lo hacen en el desempeño. En particular, se hace hincapié en el abandono reciente de las políticas aplicadas por los sistemas de compensación, basados en los insumos, en favor de sistemas basados en resultados. Estos sistemas se centran en resultados o productos educativos mensurables (como los reflejados en los resultados de las pruebas de los alumnos) a partir de los cuales se recompensan características o actividades que se supone mejoran la eficacia de la enseñanza (OCDE, 2009a). De tal modo, el capítulo ofrece un panorama de los principales elementos del diseño de políticas públicas de incentivos docentes. Las autoridades educativas de un país en particular, junto con los actores implicados,

necesitarán determinar la combinación correcta de los incentivos y estímulos monetarios y no monetarios más eficaces. Sin embargo, además de las recompensas y consecuencias para los maestros, las prácticas vigentes indican que, para que los maestros se consideren eficaces, sus alumnos deben demostrar niveles satisfactorios de crecimiento, por lo que ningún maestro debe considerarse ineficiente si sus alumnos muestran niveles satisfactorios de crecimiento (*Delaware Department of Education*, 2010). A partir de las experiencias e investigaciones internacionales, el apéndice al final del capítulo brinda igualmente consideraciones detalladas para la prueba piloto, el monitoreo y la evaluación de las políticas y programas de incentivos docentes. Este apéndice pretende ofrecer directrices sobre el diseño, la planeación y la implementación de procesos eficaces de monitoreo y evaluación.

7.1 TIPOS DE INCENTIVOS DOCENTES

Se supone que ofrecer incentivos financieros incrementa la productividad organizacional pues esto alienta a los maestros menos eficientes a buscar estrategias de instrucción más eficaces, a reducir la rotación de los maestros mejor capacitados, y a atraer a maestros que son particularmente buenos en las actividades requeridas. En la medida en que dichos incentivos generen una fuerza laboral docente más eficiente, esos incentivos tendrán impactos positivos en las prácticas de enseñanza mucho más allá de la vida operativa de un programa en particular.

Incentivos basados en insumos

Los incentivos basados en insumos son los que recompensan a los maestros por actividades que se supone mejoran la calidad de la instrucción o los resultados de los alumnos. En otras palabras, los incentivos basados en insumos tienen el propósito de hacer que los maestros prefieran realizar actividades pensadas para mejorar la eficacia, aunque no logran indicar si la eficiencia general de los maestros de verdad mejora. Esta característica ha provocado que mucha gente critique, y en algunos casos evite, este tipo de incentivos. Entre los ejemplos más comunes de incentivos basados en insumos, se encuentran los otorgados por la formación docente, el pago conforme a conocimientos y habilidades, y los escalafones profesionales. El programa Carrera Magisterial en México, antes de 2006, es un caso ilustrativo.

Incentivos basados en conocimientos y habilidades, y escalafones profesionales

Los incentivos basados en conocimientos y habilidades recompensan a los maestros por la adquisición de conocimientos y habilidades adicionales que se supone mejoran la eficiencia general de un docente. El pago con base en conocimientos y habilidades es distinto al aumento del sueldo para grados más avanzados previsto en el esquema salarial tradicional. Precisamente, los incentivos se centran en la mejora y crecimiento continuos de las habilidades y competencias docentes a lo largo de la carrera. La premisa de este pago es motivar a los maestros a que mejoren sus conocimientos y habilidades, y mejorar así (en teoría) sus prácticas de instrucción, incrementar su eficiencia y aumentar el logro de los alumnos (Odden y Kelley, 1997).

Los incentivos basados en conocimientos y habilidades pueden recompensar a los maestros en un nivel superior al que pueden aspirar gracias al incremento previsto en el esquema salarial tradicional al obtener un grado más avanzado. Otros casos son tomar cursos adicionales de desarrollo profesional, procurar una certificación dual, y completar un portafolio de enseñanza. En Singapur, además de las 100 horas de desarrollo profesional a las que tienen derecho cada año, los maestros pueden solicitar reembolsos de entre 400 a 700 dólares de Singapur por gastos relacionados con la mejora de sus conocimientos y habilidades. Los gastos permitidos son la compra de *software*, tomar cursos de formación, la suscripción a publicaciones e integrar organizaciones profesionales.

Es común que el pago basado en conocimientos y habilidades implique combinar el aumento de salario a evaluaciones y mediciones externas. Por ejemplo, en los Estados Unidos de América estos bonos suelen otorgarse a maestros que participan y completan la Certificación de la Junta Nacional de Estándares Profesionales de Enseñanza (*National Board for Professional Teaching Standards Certification*), o que aprueban otros exámenes de evaluación docente (Podgursky y Springer, 2007). Dependiendo del diseño del sistema, los maestros reciben pagos anuales durante una cierta cantidad predeterminada de años, o por lo que dure su trayectoria profesional.

Semejantes al incentivo económico basado en los conocimientos y habilidades, los escalafones profesionales crean distintas categorías, o niveles, que recompensan a los maestros con salarios más elevados. Cada nivel se asocia a un mayor dominio o competencia, y distingue a los maestros nuevos de los expertos o con postgrado. En la mayoría de los casos, los escalafones profesionales requieren que los maestros aprueben una certificación oficial o no oficial, o que asuman responsabilidades adicionales, con el fin de ascender por el escalafón. En general, los maestros deben cumplir con múltiples criterios para avanzar. Por ejemplo, en el Reino Unido la escala superior del esquema salarial incluye el pago por desempeño basado en ocho estándares de enseñanza acordados nacionalmente. Los maestros que alcanzan la parte superior de la escala salarial (a la que se llega en cinco pasos) pueden solicitar pasar el “umbral” mediante una evaluación conforme a dichos estándares, agrupados en cinco categorías: conocimiento y comprensión (un estándar), enseñanza y evaluación (tres estándares), avance de los alumnos (un estándar), eficiencia profesional más amplia (dos estándares) y características profesionales (un estándar). Se forma a los directores para completar la evaluación, validada después mediante revisores externos. En las escalas de pago superiores, en las que los maestros avanzan cada dos años mediante evaluaciones anuales, los salarios docentes son entre 12% y 25% más altos. Los escalafones profesionales ofrecen nuevos roles a los maestros mediante una retribución económica y responsabilidades adicionales conforme incrementan sus conocimientos y habilidades. Además de ofrecer posibilidades a los buenos maestros para que avancen (con lo que se les anima a continuar en servicio), los programas de escalafón profesional “contrarrestan el estancamiento al variar las responsabilidades y actividades de los maestros en cada nivel” (Cresap *et al.*, 1984, p. 22).

Compensación basada en el desempeño

En parte por las críticas hacia los incentivos basados en insumos, y también como resultado del uso cada vez más extendido de las evaluaciones estandarizadas, los planes de pago de incentivos se han enfocado cada vez más en recompensar a los maestros o grupos de maestros basándose en su desempeño o en sus resultados. Los sistemas de compensación basados en el desempeño cobraron una mayor importancia a la luz de las sólidas evidencias que mostraron que muchos sistemas basados en insumos son financieramente ineficaces. Por ejemplo, en los Estados Unidos de América, los investigadores Roza y Miller (2009) observaron que se gastaron 8.6 billones de dólares estadounidenses en pagos por postgrados, lo que representó un aumento de 3.3% de los gastos educativos totales en algunos estados, pese a la evidencia de las investigaciones que demuestra que las maestrías en materias generales no guardan relación con una mayor eficiencia docente.

Los incentivos basados en el desempeño se asocian de manera más directa al aprendizaje del alumno que los incentivos basados en acciones, y por lo general se agrupan en dos categorías: los basados en resultados educativos y los basados en procesos educativos. Si bien estos programas pueden no sólo emplear múltiples medidas para evaluar el desempeño docente, sino también incorporar elementos propios del escalafón profesional o de los planes basados en la retribución económica por conocimientos y habilidades, los resultados de los alumnos en evaluaciones estandarizadas siguen siendo indispensables para determinar a quién hay que dar los bonos.

Incentivos por resultados en el desempeño

Los incentivos por resultados en el desempeño son los que dependen de las mediciones cuantificables del logro del alumno como herramienta principal de evaluación. Estos bonos se distinguen en que la recompensa “depende de los resultados del alumno atribuidos a un maestro o grupo particular de maestros” (Podgursky y Springer, 2007). Las puntuaciones de los alumnos en evaluaciones estandarizadas, índices de graduación, índices de abandono, índices de asistencia e índices de avance de grado, son ejemplos de los tipos de mediciones posibles. Los incentivos por resultados pueden considerarse los que recompensan al producto de la enseñanza.

Cada vez más, los incentivos por resultados en el desempeño se basan en el desempeño del alumno medido en evaluaciones estandarizadas. Esto se debe en gran medida tanto a las mejoras en la elaboración e implementación de las evaluaciones estandarizadas en todos los países, y a las mejoras en los sistemas de datos longitudinales

que vinculan a alumnos y maestros en lo individual, como a una medición más compleja del desempeño de alumnos, maestros y escuelas (OCDE, 2010b). Los resultados de la evaluación del alumno también sirven para una retroalimentación y evaluación docentes. De acuerdo con la Encuesta Internacional sobre Docencia y Aprendizaje (TALIS por sus siglas en inglés) de la OCDE, de más de 70 000 maestros en 23 países, 65% de los maestros reporta que las puntuaciones en las pruebas de los alumnos fueron un criterio moderado o muy importante para la evaluación o retroalimentación que recibieron. En Bulgaria, Malasia, México y Polonia, este porcentaje rebasó por mucho el 80% (OCDE, 2009b).

En los Estados Unidos de América, la ley Ningún Niño Rezagado (*No Child Left Behind*), aprobada en 2002, dicta que cada año todos los estados deben aplicar ciertas pruebas a los alumnos del tercero al octavo grados, y una vez en el nivel superior, tanto en lectura como en matemáticas. Asimismo, cada estado debe publicar el nivel de resultados de estas evaluaciones. Como resultado de esta y otras iniciativas, muchos estados elaboraron amplios sistemas de datos del logro del alumno, incluyendo los vínculos alumno-maestro. En combinación con iniciativas de políticas que pretenden mejorar el logro de los alumnos, los programas de pago de incentivos basados en el logro de los alumnos se difundieron a gran escala en los ámbitos federal, estatal, distrital y escolar de los Estados Unidos de América.

A finales de la década de 1980, en el Reino Unido se introdujo una nueva política de rendición de cuentas escolar mediante la Ley de Reforma Educativa (*Education Reform Act*) de 1988. El sistema instituyó una serie de reformas educativas nacionales con el fin de elevar los logros. El sistema de rendición de cuentas escolar del Reino Unido evolucionó a un marco amplio y más complejo de rendición de cuentas que incluye: a los copartícipes de la mejora escolar; los requerimientos de aprendizaje continuo para maestros con más antigüedad; las recompensas y sanciones para escuelas cuyo desempeño sea considerado deficiente; la elección de la escuela; la evaluación interna y externa de procesos de instrucción y organizacionales; y un acento en el desarrollo profesional y apoyo continuos para maestros y líderes escolares.

Un avance significativo en los datos de evaluación de los alumnos es el uso de métodos estadísticos conocidos como modelos de valor agregado (MVA), como se analizó en el capítulo 5, los cuales aíslan la contribución por parte de escuelas, grupos de escuelas y, en algunos casos, maestros en lo individual, al aprendizaje de los alumnos. Aunque las puntuaciones brutas pueden proporcionar información útil sobre los logros de aprendizaje de los alumnos y reforzar los sistemas de rendición de cuentas y acciones de mejora, por sí solas las puntuaciones brutas pueden contener resultados distorsionados y sesgados del desempeño escolar (véase un análisis más detallado en el capítulo 5). Sin embargo, con modelos de valor agregado, los investigadores tienen la posibilidad de medir con mayor precisión el crecimiento de los alumnos, y en qué grado dicho crecimiento es atribuible a escuelas y maestros, independientemente de los antecedentes socioeconómicos y otras características familiares de los alumnos. Este tipo de información brinda una idea mucho más profunda de la eficacia de las escuelas y, de manera indirecta, de los maestros. Aunque los modelos de valor agregado son la vanguardia en la innovación e iniciativas de políticas educativas en sólo unos cuantos países (OCDE, 2010b), estos ofrecen una opción importante a los sistemas que requieren mediciones más precisas y exactas del desempeño escolar y eficiencia docente de las que se obtienen a partir de puntuaciones brutas de pruebas, u otras medidas, que puedan sufrir sesgos y problemas de validez.

Ejemplos de incentivos por resultados de desempeño

En los Estados Unidos de América, el Fondo de Incentivos Docentes (*Teacher Incentive Fund*, TIF por sus siglas en inglés) ofrece becas federales a maestros y directores de escuelas necesitadas para apoyar el desarrollo y la implementación de sistemas de compensación basados en el desempeño. Este Fondo ha otorgado becas desde 2007 a unos 95 beneficiarios, incluyendo a 62 en septiembre de 2010. Diseñados e implementados a nivel local, todos los programas deben incluir el logro académico de los alumnos, así como evaluaciones en el salón de clases entre otros factores, para determinar a quiénes se otorgan los incentivos. En un análisis de 2010 de los primeros 33 beneficiarios del Fondo, Heybourn *et al.* observaron, como se esperaba, que los empleados

podieran recibir compensaciones con base en el desempeño de los alumnos, con un poco más de la mitad que dependían de alguna forma de medición de valor agregado. Con un financiamiento original de 99 millones de dólares estadounidenses en 2006, el Fondo recibe hoy en día 400 millones de dólares estadounidenses, y las becas comienzan a partir de 500 000 dólares estadounidenses. Además, el Departamento de Educación estadounidense obtuvo 950 millones de dólares estadounidenses para 2011 para destinarlos a un nuevo Fondo de Innovación de Maestros y Líderes (*Teacher and Leader Innovation Fund*), el cual apoyaría la elaboración e implementación de enfoques orientados al desempeño para seleccionar, retener y recompensar a los educadores más eficientes (Springer, 2010).

Además de los programas federales, varios estados y distritos implementaron programas de incentivos basados, al menos en parte, en el logro de los alumnos. En Texas, el Programa de Premios del Gobernador a la Excelencia de los Educadores (*Governor's Educator Excellence Award Program*) que se subdivide en tres programas, fue el sistema de financiamiento estatal más grande, con un fondo cercano a los 250 millones de dólares estadounidenses. El primero de estos programas, la Beca del Gobernador a la Excelencia del Educador (*Governor's Educator Excellence Grant*), proporcionó becas de tres años a 99 escuelas de altos méritos, ubicadas entre el tercio superior de escuelas que atendían a alumnos desfavorecidos. Las escuelas participantes diseñaron sus propios planes, pero se les pidió que incluyesen puntuaciones del logro de los alumnos y mediciones de la colaboración docente. Para el componente de logro del alumno, los premios se basaron sobre todo en el desempeño individual de los maestros y en los niveles de aprovechamiento de los alumnos, en lugar de hacerlo en sus puntuaciones de crecimiento. Durante el transcurso de los programas, los maestros participantes recibieron premios que iban de menos de 100 a más de 10 000 dólares estadounidenses, con un promedio por premio de unos 2 300 dólares estadounidenses en los tres años (Springer *et al.*, 2009).

El segundo programa de Texas, Beca a la Excelencia del Educador de Texas (*Texas Educator Excellence Grant*), otorgó más de 1 000 becas por año durante dos años a escuelas de mejor desempeño que atienden a una proporción elevada de alumnos en desventaja. Como sucede con el Programa de Premios del Gobernador a la Excelencia de los Educadores, los planes se diseñaron al nivel de la escuela y el distrito, manteniendo la necesidad de medir el logro de los alumnos y la colaboración docente. En general, las escuelas que participan en el Programa de Premios del Gobernador a la Excelencia de los Educadores emplean puntuaciones del logro de los alumnos, en lugar de puntuaciones del crecimiento de los alumnos, estableciendo así la medición del desempeño. Es interesante observar que en los dos años que duró el programa, cerca del 90% de las escuelas dependió sólo de evaluaciones estandarizadas estatales para medir los resultados del alumno, pese a contar con diversas fuentes alternas. El promedio de los bonos en el primer año del programa fue de 1 982 dólares estadounidenses, mientras que en el segundo, los premios promediaron 2 094 dólares.

El Programa de Premios Distritales a la Excelencia Docente (*District Awards for Teacher Excellence Program*), el tercer programa de Texas, proporcionó fondos a nivel distrital de manera abierta a todos los distritos del estado. Como con los dos primeros programas, los planes se diseñaron localmente y se requirió que incluyesen mediciones del logro de los alumnos. En total, participaron 203 entidades en el programa de Premios Distritales, lo cual representa 16% de todos los distritos elegidos (Springer *et al.*, 2009). De los que eligieron no participar, a una clara mayoría le preocupó el impacto del programa en la cultura de la escuela y las asociaciones profesionales (Springer, 2010).

En la Ciudad de Nueva York, el Programa de Pago de Bonos al Desempeño Basado en la Escuela (*School-Based Performance Bonus Pay Programme*), implementado en 200 escuelas públicas de primaria y secundaria durante el año lectivo 2007-2008, otorgó bonos a escuelas de hasta 3 000 dólares estadounidenses por cada miembro de un sindicato de tiempo completo que trabajó en la escuela para cumplir con objetivos de desempeño definidos por el sistema de rendición de cuentas de la ciudad. Experimento de aplicación aleatoria, este programa es único ya que el resultado de desempeño de interés es el resultado de la escuela en su conjunto. De hecho, la medición y distribución de premios a nivel escolar fueron predominantes para garantizar el apoyo

de la asociación magisterial local. Además de supervisar las puntuaciones del logro de los alumnos, el distrito recopila datos mediante encuestas sobre la percepción que tienen alumnos, padres y maestros del ambiente de aprendizaje de la escuela, incluyendo preguntas sobre las expectativas académicas, la comunicación, el compromiso y la seguridad. Asimismo, equipos de educadores experimentados efectúan visitas de campo de dos a tres días para revisar la calidad del programa institucional y de instrucción de una escuela. La encuesta del ambiente de aprendizaje de la escuela y las revisiones de calidad externas ofrecen los medios para estudiar el efecto causal de este programa en resultados intermedios. A diferencia de la gran mayoría de los programas de pago por desempeño, éste se diseñó como experimento aleatorio, lo que permite a los investigadores medir su impacto en el aprendizaje de los alumnos.

Los resultados de las investigaciones sobre el efecto de estos programas en el logro de los alumnos aún están por verse. Como con los programas anteriores basados en insumos, se necesitan más investigaciones para evaluar la eficacia de estos programas como mecanismo para la mejora del aprendizaje de los alumnos, en el corto y el largo plazo. Hay extensos y numerosos estudios en curso, o que se encuentran en su etapa inicial, incluyendo una evaluación aleatoria del programa Fondo de Incentivos Magisteriales en los Estados Unidos de América (Springer, 2010). Más adelante, será importante retomar las lecciones de las investigaciones para afinar y mejorar estos programas.

Otras medidas del aprendizaje de los alumnos

Como se observó en los capítulos anteriores, las puntuaciones de las pruebas no son la única medida de resultados del desempeño de los alumnos. Los índices de asistencia, los índices de graduación, los índices de abandono y los portafolios de los alumnos, son otros ejemplos comunes de criterios que se pueden incluir en los planes de incentivos. En Austin, Texas, los Objetivos de Aprendizaje de los Alumnos (*Student Learning Objectives*, SLO por sus siglas en inglés) son la medición principal del logro docente individual utilizada para calcular el pago por el desempeño, conocida como el programa REACH del Distrito Escolar Independiente de Austin (*Austin Independent School District*, AISD por sus siglas en inglés) (Springer, 2010; AISD, 2010). Los SLO son metas de instrucción diseñadas por los maestros y basadas en los datos. Mediante el examen de los datos de logro de los alumnos, los maestros trabajan con el director de su escuela para elaborar dos SLO basados en las necesidades de los alumnos. Al menos un SLO debe estar destinado a la clase del maestro en su conjunto, mientras que el segundo SLO se centra en un subconjunto particular de alumnos. Los SLO deben basarse en estándares establecidos por el gobierno estatal, abordar necesidades del salón de clases, alinearse con las metas del Plan de Mejora del Campus (*Campus Improvement Plan*) y satisfacer los estándares de rigor tanto de desempeño como de evaluación. Además, cada SLO debe recibir no sólo la aprobación del director del maestro, sino también la del personal del programa REACH AISD, para garantizar que los objetivos sean adecuados y rigurosos. Al final del año lectivo se evalúa a los alumnos para determinar si se cumplió con el objetivo de desempeño. Los maestros llegan a percibir hasta 3 000 dólares estadounidenses por cumplir ambas metas SLO, y los directores hasta 4 500 dólares estadounidenses por facilitar el proceso. Un sistema casi idéntico llamado Objetivos de Crecimiento de los Alumnos (*Student Growth Objectives*, SGO por sus siglas en inglés) ofrece una variable de resultados del logro de los alumnos en Denver, en el programa de pago por desempeño (ProComp por sus siglas en inglés) de Colorado (Springer, 2010).

Si bien los SLO y los SGO comprometen a los maestros a usar datos para orientar la instrucción y las metas curriculares, también carecen de la objetividad de una evaluación estandarizada. De hecho, una evaluación del programa REACH AISD reveló que 83% de los maestros en el primer año, y 81% en el segundo, cumplieron al menos con uno de sus SLO (Cornetto *et al.*, 2010). Este elevado nivel de logro plantea dudas sobre el rigor general de las metas y si representan adecuada, o verdaderamente, la eficiencia docente. Esto subraya la importancia de garantizar que las metas y medidas sean tan rigurosas como objetivas al diseñar un programa de incentivos.

Incentivos de procesos

A diferencia de los incentivos basados en el desempeño, los incentivos de los procesos de desempeño se centran en las prácticas de enseñanza y en las habilidades del maestro en el salón de clases. Por ejemplo, en algunos estados de Alemania los maestros tienen la posibilidad de avanzar en los niveles salariales más rápido de lo normal cuando se someten a evaluaciones de desempeño basadas en observaciones en el salón de clases. En Baden-Wurtemberg, 10% de los maestros de todas las escuelas pueden avanzar una etapa salarial a partir de observaciones externas, mientras que en Renania del Norte-Westfalia, los maestros sólo pueden avanzar en los niveles salariales con base en evaluaciones de desempeño.

En un esfuerzo por establecer medidas objetivas, válidas y confiables del desempeño docente sin incluir datos del desempeño de los alumnos, los sistemas procuran incorporar medidas de eficiencia docente que no reciban una influencia negativa de sindicatos magisteriales, lealtades locales u otros intereses (Buck y Greene, 2010). Por ejemplo, las sesiones de evaluación docente en Singapur incluyen una planeación extensa, una reunión para fijar las metas al principio del año lectivo, una revisión y reunión a mitad del año, y una evaluación final con recomendaciones y retroalimentación por parte de maestros con más antigüedad y de encargados de departamentos o áreas temáticas que trabajaron con los maestros. Los maestros que consiguen altas clasificaciones en su evaluación de desempeño pueden ser postulados para recibir bonos equivalentes de entre uno y tres meses de su salario.

En Suiza, el sistema de calificación de salario efectivo examina a los maestros a la mitad de su trayectoria. En este caso, un equipo de profesionales especialmente formados observa y entrevista al maestro, y prepara un reporte completo del enfoque pedagógico del docente, gracias al cual los maestros exitosos pueden ser postulados para recibir aumentos de pago de 1% a 4% durante los cuatro años posteriores.

Las evaluaciones de procesos de desempeño ofrecen una buena aproximación a las prácticas del maestro en el salón de clases, y son benéficas para brindar retroalimentación personalizada al docente. Sin embargo, como se mencionó anteriormente, es crítico vincular estas observaciones con mediciones objetivas del aprendizaje de los alumnos. En 2010, la Fundación Gates donó 45 millones de dólares estadounidenses para lanzar el Proyecto de Medidas de Enseñanza Eficaz (*Measures of Effective Teaching Project*) (Springer, 2010). El proyecto de dos años, a cargo de investigadores educativos independientes, “pretende idear y desarrollar un conjunto de medidas que funcionen de manera conjunta para obtener un indicador más completo del impacto de un maestro en el logro de los alumnos” (Gates Foundation, 2010). Los investigadores trabajan con 3 000 maestros en siete distritos escolares en los Estados Unidos de América, y recopilan información como: la retroalimentación de los alumnos mediante encuestas; muestras de trabajos de los alumnos; evaluaciones complementarias de los alumnos; lecciones videograbadas en el salón de clases; reflexiones de los maestros sobre sus clases videograbadas; evaluación de la capacidad de los maestros para reconocer y diagnosticar problemas de los alumnos; y encuestas a los maestros sobre las condiciones laborales. El propósito del estudio no es investigar el diseño de los incentivos sino, por el contrario, determinar las prácticas de instrucción reales de los maestros eficientes. Los resultados de este amplio estudio pueden brindar información invaluable para los sistemas internacionales de evaluación, incluyendo criterios más rigurosos con los cuales vincular el pago del maestro a su desempeño.

Reformas orientadas al mercado

La compensación orientada al mercado recompensa a los maestros según los factores del mercado, tales como la ubicación y la materia que se enseña, y suele ser una respuesta a consideraciones más amplias de la oferta y la demanda al interior de la escuela, el distrito o la escuela en particular. Las formas más comunes de compensación orientada al mercado recompensan a los maestros asignados a una escuela con dificultades para contratar personal o a los que enseñen una materia difícil de cubrir.

Un mecanismo común para otorgar una compensación orientada al mercado es el pago de bonos de selección y retención. La remuneración de selección se otorga a los nuevos maestros en una escuela y sirve como herramienta para motivar al maestro a aceptar un puesto docente en esa escuela (por lo general una escuela con dificultades

para dotarse de personal). La remuneración de selección se distribuye a menudo sin una relación directa con la experiencia en la enseñanza (esto es, el maestro en cuestión puede ser uno con experiencia que fue reasignado o un principiante), y suele hacerse a nivel escolar o distrital de manera individual. En algunos sitios, la remuneración de selección se otorga durante varios años, después de que un maestro haya aceptado un nuevo puesto. La remuneración de retención se distribuye a maestros que continúan trabajando en su puesto y/o escuela después de las vacaciones de verano. Por lo general, se considera que la remuneración de retención mejora el logro de los alumnos en escuelas con dificultades para dotarse de personal, al disminuir en ellas la rotación continua y la emblemática abundancia de maestros principiantes.

En Singapur, el programa Continuidad, Experiencia y Compromiso (*Continuity, Experience and Commitment*, CONNECT por sus siglas en inglés) ofrece pagos de retención a maestros jubilados, además del acceso al seguro social disponible para todos los maestros. Los pagos van de 4 200 a 6 200 dólares de Singapur al año para maestros con una experiencia de 1 a 15 años, y de 3 200 dólares de Singapur al año para maestros con mayor experiencia. Los fondos también se pueden retirar cada tres o cada cinco años para ayudar a fomentar la retención hasta la jubilación. En Queensland, Australia, además de los incentivos de reclutamiento, se ofrece a los maestros 5 000 dólares australianos al año por permanecer en su puesto docente.

Materias de difícil cobertura

Los incentivos para materias de difícil cobertura son los que se ofrecen a maestros, o a futuros maestros, calificados para enseñar materias que experimentan una escasez, y que se basan en las necesidades de las escuelas, del distrito o del estado. Un reporte de la OCDE de 2005 reveló que más de 30% de alumnos en los países participantes asistía a escuelas con escasez de maestros calificados en lengua extranjera, matemáticas, ciencias y tecnología (OCDE, 2005). De manera más general, a nivel internacional las materias de ciencias, tecnología y matemáticas padecen más escasez, pues un menor número de alumnos universitarios de estas áreas acceden a la profesión. Debido a las significativas diferencias salariales entre los trabajadores de los sectores público y privado en estas áreas, los incentivos son un mecanismo popular para cubrir esta brecha. En los Estados Unidos de América, los incentivos para materias de difícil cobertura se ofrecen con más frecuencia para matemáticas, ciencias y maestros de educación especial, con remuneraciones que suelen ir de los 1 500 a los 3 000 dólares estadounidenses. Además, muchos distritos y estados ofrecen asistencia en colegiaturas y condonación de préstamos para maestros de áreas designadas en escasez. En Utah, los maestros calificados de matemáticas y ciencias reciben un bono de contratación de 5 000 dólares estadounidenses por acceder a permanecer en el distrito durante al menos cuatro años. En la Ciudad de Nueva York, donde los costos de vivienda exceden el promedio nacional, los maestros calificados de matemáticas, ciencias y educación especial pueden obtener ayuda para la vivienda de hasta 15 000 dólares estadounidenses para gastos de mudanza, enganches o fianzas para alquiler, y una remuneración de vivienda de 400 dólares durante dos años (Springer, 2010).

En Inglaterra y Gales, los programas de condonación de préstamos reembolsan durante diez años hasta 16 000 libras esterlinas de gastos de colegiaturas a los maestros calificados de matemáticas, ciencias, educación especial y tecnología. El programa Saludo Dorado (*Golden Hello*) ofrece 4 000 libras esterlinas adicionales a maestros que completan su primer año de enseñanza en un área de escasez. En Australia e Irlanda se ofrecen incentivos a maestros de idiomas indígenas, mientras que Bruselas ofrece incentivos a maestros de instrucción del idioma francés (OCDE, 2009a).

Escuelas con dificultades para dotarse de personal

Los incentivos para enseñar en escuelas con dificultades para dotarse de personal se dirigen a escuelas, distritos o regiones geográficas que experimentan una insuficiencia crónica de maestros calificados en todas o casi todas las materias. Por lo general, estas escuelas se ubican en áreas urbanas o rurales remotas, y suelen atender a una alta proporción de alumnos económicamente desfavorecidos, de minorías o de bajo desempeño, y se caracterizan no sólo por tener un elevado índice de rotación de personal entre los maestros, sino también docentes relativamente faltos de experiencia.

Muchos países y estados ofrecen incentivos para seleccionar y conservar a maestros en regiones rurales remotas. En Australia, el Esquema de Incentivos de Áreas Remotas de Queensland (*Queensland Remote Area Incentive Scheme*) (Sclafani y Tucker, 2006) ofrece a los maestros rurales 5 000 dólares australianos para viajar desde áreas metropolitanas, además de incentivos de retención de hasta 5 000 dólares. En Chile, los maestros que trabajan en áreas de aislamiento geográfico, pobreza extrema o de difícil acceso, reciben incentivos de hasta un 30% de su salario anual. En otros países, como Corea del Sur, China y alguna vez Bolivia, se requiere que los maestros pasen un tiempo enseñando en áreas remotas para ser elegibles para ascensos.

Aunque contar con suficientes docentes en ambientes rurales remotos es un reto en algunos países, otros países y estados experimentan lo opuesto: carencias extremas en áreas urbanas. Por ejemplo, el Incentivo de Sobresueldo por Contratación (*Staffing Incentive Allowance*) en Nueva Zelanda proporciona 1 000 dólares neozelandeses a los maestros que trabajen tres años en escuelas urbanas designadas de alta prioridad. De igual modo, en los Estados Unidos de América, los maestros en Los Ángeles, California, en Charlotte, Carolina del Norte, y en la Ciudad de Nueva York, Nueva York, fueron elegibles para recibir incentivos de 2 000 a 3 000 dólares estadounidenses por trabajar en escuelas de alta prioridad.

Más allá de los incentivos monetarios

Los programas de incentivos se diseñan para satisfacer necesidades locales, regionales o nacionales específicas, y adoptan diversas formas, como ofrecer incentivos tanto financieros como no financieros. De hecho, Vegas y Umansky (2005) identifican diversos incentivos posibles para los docentes, por ejemplo, el reconocimiento y el prestigio, la gratificación de motivación intrínseca, la estabilidad laboral, condiciones laborales con recursos y apoyo adecuados, etc. De igual modo, Darling-Hammond (1997) observó una fuerte asociación entre las opiniones de los maestros sobre la calidad del apoyo que recibían de la administración de su escuela y sus planes para conservar su puesto de enseñanza. Darling-Hammond encontró también que la disponibilidad de recursos y la capacidad de los maestros para hacerse escuchar en las decisiones de la escuela estaban ligadas a sus planes de quedarse en el establecimiento. De igual manera, se observó que una falta de ambientes laborales positivos contribuía a índices elevados de desgaste en escuelas de áreas de gran pobreza y escuelas con grandes poblaciones de alumnos de minorías (Loeb *et al.*, 2005).

En el Reino Unido, el Programa de Enseñanza Rápida (*Fast Track Teaching Programme*) ofreció incentivos a nuevos maestros prometedores que manifestaban un liderazgo potencial. Este programa se diseñó para identificar a líderes escolares en etapas tempranas de su carrera, y los maestros en este programa son elegibles para recibir 2 000 libras esterlinas al año, así como actividades de asesoría, tutoría y desarrollo de liderazgo. También en Inglaterra, el programa Elevación de Estándares y Disminución de Carga de Trabajo –Acuerdo Nacional (*Raising Standards and Tackling Workload– A National Agreement*), pretende abordar la insatisfacción de los maestros respecto a las condiciones laborales. El programa tiene la intención de mejorar la satisfacción y la retención generales de docentes al reducir las horas del contrato del maestro, brindar garantías de tiempo de planeación, reducir los trámites burocráticos y proporcionar personal de apoyo diverso, como personal administrativo, de oficina, técnico y para el salón de clases. El programa también creó nuevas trayectorias de carrera para los docentes (Springer, 2010).

En Singapur, los programas ofrecen a los maestros incentivos tanto financieros como no financieros. Por ejemplo, se permite a los maestros acumular un mes de paga completa por un servicio de medio tiempo, o la mitad de la paga por estudios de tiempo completo por cada año de servicio, hasta por 12 meses. Después, los maestros tienen la posibilidad de dedicar su tiempo sin importar cómo siempre y cuando beneficie al sistema educativo, como, por ejemplo, enseñar en una escuela privada o extranjera, o trabajar en el sector privado en un área relacionada con su materia de enseñanza. De manera semejante, los planes de salario diferido para licencias en Canadá y algunas partes de los Estados Unidos de América permiten a los maestros diferir una parte de su salario vigente para usarla cuando soliciten una licencia más adelante. Los programas en Noruega,

Alemania y los Países Bajos permiten a los maestros reducir su carga de trabajo de cursos con una pequeña reducción de salario (OCDE, 2009a).

Si bien no hay un diseño estándar para estos programas, la intención es mejorar las condiciones laborales generales de los maestros, con lo cual teóricamente debería mejorar su eficiencia general y su índice de retención. Milanowski (2007) observó que los futuros maestros se preocupaban más por las características de su asignación docente, incluyendo el apoyo de los directores, la disponibilidad de programas de inducción y flexibilidad curricular en la escuela, que por los aumentos de salario.

La importancia de evaluar las evaluaciones

No se puede exagerar la necesidad de evaluar integralmente los programas de evaluación e incentivos docentes. Si bien las secciones precedentes ofrecen un panorama de la diversidad de programas de incentivos que utilizan los resultados del logro de los alumnos, es necesario contar con evaluaciones rigurosas para garantizar tanto la eficacia de los programas más recientes, como su rentabilidad (Springer, 2010). En una evaluación del programa Beca a la Excelencia del Educador de Texas (*Texas Educator Excellence Grant*), Springer *et al.* (2009) descubrieron que, aunque las actitudes y comportamiento del personal escolar, el ambiente de la escuela y la rotación de maestros reflejaron la influencia del programa, la evidencia sugiere que el programa no tuvo un efecto fuerte y sistemático para aumentar el logro de los alumnos. Tampoco se dieron asociaciones consistentes entre las características del diseño del programa y las ganancias del logro de los alumnos. Por otra parte, un estudio reciente en el Reino Unido reveló que los incentivos para el desempeño mejoraron el aprendizaje de los alumnos en un promedio de 40% por grado, medido mediante pruebas estandarizadas (Atkinson *et al.*, 2009). Sin embargo, este estudio se basó en datos que vinculan a los alumnos con maestros en lo individual al paso del tiempo; en muchos países esto no puede hacerse con los sistemas de datos actuales, aunque hay acciones en curso para lograrlo (OCDE, 2010b). De tal modo, las autoridades educativas y los actores implicados necesitarán elaborar programas de incentivos docentes específicos, tomando en cuenta de manera cuidadosa el contexto, las limitaciones y las oportunidades locales, como se analizó en el capítulo 2.

Consideraciones de diseño

Si bien es importante determinar el tipo óptimo de programas de incentivos (basado en insumos, en el desempeño o un híbrido), reviste igual importancia el diseño del programa. Por ejemplo, ¿los premios deben estar dirigidos a cada uno de los maestros, al equipo o a la escuela? ¿Qué umbrales de desempeño mínimo se deben alcanzar para obtener un premio, y cuáles son las clasificaciones adecuadas para los premios por incentivos? ¿Todos los maestros deben ser elegibles para el premio, o sólo los que se desempeñen mejor? La forma de estructurar estos programas para responder estas preguntas tiene un impacto significativo en la eficacia del programa en su conjunto. La siguiente sección identifica las características claves de diseño al elaborar un programa de incentivos.

Estructura de los incentivos

La estructura de los incentivos se refiere al esquema, o mecanismo, mediante el cual se asignan los premios en un sistema de pago por desempeño. En algunos casos, sólo una cantidad o proporción limitada de maestros puede ganar un premio, mientras que en otros, todo maestro que cumpla con un estándar de desempeño predeterminado recibe un premio. Las dos formas principales de estructurar los incentivos son *torneos de jerarquización* y *contratos de desempeño fijo*.

Los *torneos de jerarquización* son estructuras de incentivos en donde una proporción fija de maestros, equipos o escuelas, gana un premio con base en su jerarquía de desempeño respecto a la jerarquía de otros maestros, equipos o escuelas. La característica clave que distingue a los torneos de otras estructuras de pago de incentivos es que la compensación depende de un desempeño relativo en lugar de un desempeño absoluto.

Al fijar por adelantado la cantidad total de individuos o equipos que pueden ganar los premios, los torneos de jerarquización reducen el potencial de una exposición financiera incierta. Por otra parte, al poner a competir a los maestros entre sí, es más probable que se cree una competencia entre las escuelas y se perturbe la cultura colaborativa de la enseñanza. Más aun, si el estándar de desempeño no es claro o se percibe como inalcanzable, quienes tengan un desempeño bajo tal vez no se interesen en los incentivos.

Un *contrato de desempeño fijo* define el estándar de desempeño que deben alcanzar los maestros, equipos o escuelas, para ganar un premio. A diferencia de los torneos de jerarquización, todo individuo que cumpla con el estándar predeterminado de desempeño es elegible para recibir los beneficios, independientemente del desempeño de los demás. La exposición financiera potencial con los contratos de desempeño fijo es significativa, pero la estructura suele ser más atractiva para los maestros y asociaciones magisteriales, pues no colocan a los maestros en competencia directa entre sí.

Unidad de rendición de cuentas

Como se analiza en los capítulos anteriores, la unidad de rendición de cuentas se refiere a la entidad responsable de un producto o servicio mensurable. El desempeño en esa dimensión mensurable determina la elegibilidad para los bonos. La unidad de rendición de cuentas se define de múltiples maneras, incluyendo a *cada* maestro, a un equipo de maestros por grado-año o departamento, a todos los empleados de una *escuela* o a alguna combinación de lo anterior (*híbrido*).

Una unidad *individual* de rendición de cuentas se refiere a una política de pago de incentivos, en la cual el desempeño individual del maestro determina la elegibilidad para obtener un premio. El *equipo* como unidad de rendición de cuentas se refiere a que la elegibilidad para el premio es el producto del desempeño agregado entre los miembros de un grupo, cuyo tamaño varía de dos empleados hasta todo el personal de una escuela. Los equipos pueden incluir equipos de grado-año y departamentos por disciplina o interdisciplinarios. La *escuela* como unidad de rendición de cuentas se refiere a que la elegibilidad para los premios es el producto del desempeño agregado entre todos los empleados de una escuela.

Aunque los premios individuales crean la conexión más fuerte entre las variaciones del monto del premio y el desempeño, y ayudan a contrarrestar los problemas de parasitismo que pueden ser sintomáticos de los premios basados en equipos o escuelas, resulta más difícil identificar cuándo una enseñanza a nivel individual es más eficiente que a un nivel de equipo o de escuela. Esto es así sobre todo porque algunos resultados de pruebas estandarizadas se prestan menos a atribuirse a un solo maestro (*i.e.* si varios maestros enseñan lectura y lengua). Sin embargo, cuando se distribuyen los premios a nivel de equipos o escuelas, esta atribución es menos problemática. Los incentivos para equipos y escuelas promueven por lo general una mayor cohesión social y un compañerismo profesional, además de sentimientos de justicia, pues los trabajadores se reúnen en torno a metas compartidas. Como la conexión entre premios y desempeño es más débil en los programas basados en equipos y escuelas, es importante considerar estas ponderaciones a la luz de las metas establecidas del programa.

Estándares y umbrales de desempeño

Los estándares y umbrales de desempeño determinan el nivel requerido de desempeño que garantice la entrega de un bono para un maestro y/o equipo de maestros. Estos componentes de diseño afectan la cantidad probable de maestros que pueden obtener un bono, así como el estándar que se debe alcanzar para obtener un premio. Debido a la cantidad de modelos disponibles, y a las complejidades involucradas al intentar poner en práctica distintos modelos de estándares y umbrales de desempeño, nos centraremos en dos modelos: el *modelo lineal limitado* y el *modelo limitado de función por pasos*.

Un *modelo lineal limitado* es aquel en que se fija un mínimo de desempeño asociado a un bono. Más allá de este punto, existe una relación lineal positiva entre el aumento de desempeño y el monto del premio. En algunos "topes" de bonos, mayores incrementos del desempeño no se asocian a un mayor monto del bono.

El espacio entre los umbrales mínimo y máximo se conoce como la zona de incentivo. El modelo permite el avance en cualquier punto en la zona de incentivo para incrementar el monto del premio, aunque sigue limitando el tamaño del bono de quienes tienen un bajo desempeño y reduce el riesgo financiero de quienes tengan un desempeño excesivamente elevado. De manera semejante al modelo lineal limitado, el *modelo limitado de función por pasos* emplea umbrales mínimos y máximos para determinar la zona de incentivo. Sin embargo, en este modelo la zona de incentivo consta de distintos umbrales de desempeño, cada uno de los cuales se asocia a un monto particular del bono. Aumentar los montos de los premios sólo se logra al alcanzar los diversos umbrales dentro de la zona de incentivo.

Tanto el modelo lineal limitado, como el limitado de función de pasos, ofrecen no sólo los beneficios de mantener un vínculo claro entre las metas de desempeño y los pagos de premios, sino que reducen también al mismo tiempo los riesgos financieros. Sin embargo, estas ventajas no liberan a quienes diseñan un sistema de pagos de incentivos de la necesidad de considerar cuál es el equilibrio correcto entre motivar el desempeño y mantener asequibles las metas de desempeño. Los estándares y umbrales de desempeño deben considerar tanto la asequibilidad (es decir, el punto en que los maestros promedio cumplen los objetivos esperados) que motiva al comportamiento, como la exclusividad que puede discriminar el desempeño.

Monto y distribución de premios por desempeño

El monto del bono o nivel del pago se refiere a la cantidad del bono que puede obtener una escuela, un equipo de maestros o cada maestro de manera individual. La distribución se refiere a las directrices que determinan la proporción de maestros que reciben un bono y cómo varían los bonos entre estos. Si bien no existe una guía clara en cuanto al tamaño óptimo de un bono dentro de un programa de pago de incentivos docentes, varios estudios sugieren que, a menudo, los premios para los maestros han sido tan pequeños que se puso en riesgo su valor motivacional (Malen, 1999; Chamberlin *et al.*, 2002; Heinrich y Marschke, 2007; Taylor *et al.*, 2009). Por ejemplo, en dos programas de Texas, las cantidades mínimas de los bonos fueron de apenas 0.4% y 1.5% del salario mensual de los maestros. Del mismo modo, los bonos del Incentivo Colectivo a Escuelas de Bolivia comenzaron con un 5% del salario mensual, mientras que el Sistema Nacional de Evaluación de Desempeño de los Establecimientos Educativos de Chile comenzó con un 4.7% (OCDE, 2009a; Springer, 2010).

No obstante, la magnitud de los bonos máximos varía. El programa Carrera Magisterial de México ofrece bonos de hasta 224% del salario mensual de un maestro. De igual manera, el Experimento POINT en Nashville, Tennessee, y el Programa de Premios del Gobernador a la Excelencia de los Educadores, en Texas, ofrecieron premios máximos equivalentes a 270% del salario mensual (Springer, 2010). Tales niveles de premios máximos sugieren que los incrementos son una respuesta a las acciones anteriores de reformas que brindaban montos poco atractivos de los premios, así como el resultado de un mayor interés en averiguar si los maestros responderían a premios sustanciales, aunque el sistema de incentivos carezca de una variedad completa de medidas.

Los sistemas de distribución de bonos determinan la igualdad con que un programa de pagos de incentivos distribuye los premios a los empleados elegibles. Un plan de distribución igualitario distribuye el incentivo monetario de manera amplia, en contraste con los planes que otorgan grandes sumas de dinero a menos escuelas, equipos de maestros o maestros en lo individual. Sin embargo, no hay una guía clara que indique si un programa de pagos de incentivos debe premiar a un número relativamente menor de maestros en una escuela, o recompensar a un número menor de maestros con un premio relativamente grande (Taylor *et al.*, 2009). Los defensores de los planes de recompensas individuales sostienen que ayudan: a crear un sistema de méritos capaz de retener a quienes se desempeñen mejor en una organización; a atraer a personas con talento semejante a largo plazo; a enviar una señal clara a quienes tienen un bajo desempeño para que mejoren o se vayan; y que son más rentables (Milgrom y Roberts, 1992; Zenger y Marshall, 1992; Ehrenberg y Smith, 1994; Pfeffer y Langston, 1993).

Al mismo tiempo, un creciente número de investigaciones sugiere que las distribuciones igualitarias promueven la cooperación y el desempeño de grupo, lo cual es básico en las organizaciones participativas. Sin embargo, Milgrom y Roberts (1992) sugirieron que una mayor dispersión de pagos puede elevar el desempeño de quienes tienen un menor desempeño. La falta de evidencias sólidas sobre qué estrategia es más eficaz para determinar los niveles de los bonos, subraya aún más la importancia de los programas piloto para determinar los umbrales y componentes más eficaces para la localidad donde se implementa un programa en particular.

Frecuencia de los pagos

La frecuencia de los pagos es el índice de la distribución de premios y el intervalo que hay entre la evaluación de la actividad deseada y la distribución de los premios. En el sector educativo, la mayoría de los sistemas de pagos de incentivos distribuye los premios anualmente, por lo general de acuerdo con el año lectivo. Esto suele deberse, al menos en los sistemas de incentivos que dependen de las puntuaciones de las pruebas de los alumnos, al calendario anual de las pruebas. En la práctica, a menudo pasa mucho tiempo entre el desempeño y el pago respectivo. Después de aplicar las pruebas a los alumnos, se necesitan meses para analizar los resultados, calcular los pagos y distribuir los premios entre los maestros. Lo común es que los maestros reciban los premios a mitad del año lectivo posterior al año en que se evaluó el desempeño.

La literatura sobre psicología sugiere que los incentivos son más eficaces cuando se recompensa el desempeño en el menor lapso de tiempo posible entre la acción buscada y la distribución de los premios (Skinner, 1981). Demasiado tiempo entre una y otra situación reduce la asociación del maestro entre lo que hizo bien y el momento en que recibe el premio. El refuerzo consistente es una poderosa fuerza de motivación, por lo que desde esta perspectiva se sugiere que lo mejor sea entregar varios pagos de premios a lo largo del año. Sin embargo, pagar incentivos más de una vez al año exige mayores recursos para supervisar de manera continua el desempeño.

Como no hay un modelo de incentivos para maestros en servicio que se pueda transplantar completamente de un contexto a otro, se deben analizar y tomar en cuenta las mejores prácticas pertinentes de los ejemplos internacionales teniendo en mente las limitaciones y oportunidades locales (como se analizó en el capítulo 2). Además, también es importante aplicar programas piloto de los esquemas de incentivos antes de su aplicación completa, esto para asegurar la eficacia y el diseño adecuado de las políticas públicas (véase el Apéndice al final de este capítulo). Como la mayoría de los países de la OCDE descentralizaron sus sistemas educativos y transfirieron en diverso grado la autoridad a órganos locales, las políticas públicas de incentivos para los maestros deben abordar aspectos referentes a la ubicación de la autoridad correspondiente. La siguiente sección presenta una breve panorámica de la importancia de este punto para los incentivos de maestros en servicio.

7.2 DIRECTRICES NACIONALES E IMPLEMENTACIÓN LOCAL: IMPORTANCIA DEL EQUILIBRIO ADECUADO

Durante décadas, la descentralización del sector educativo ha sido parte integral de las reformas políticas más amplias en varios países (*i.e.* en América Latina, India y China). Los gobiernos que se esfuerzan por aumentar la cobertura, eficiencia y calidad del sector público, incluyendo la educación, han buscado un equilibrio adecuado entre la autoridad encargada de tomar decisiones y el reparto de funciones entre los distintos gobiernos central, local y regional: provincial (Argentina), estatal (México, los Estados Unidos de América), municipal (Chile), distrital (Estados Unidos de América) o una combinación de los anteriores (Brasil y los Estados Unidos de América) (OCDE, 2007; The World Bank, 2004b). Estas reformas produjeron diversos acuerdos institucionales con la distribución de funciones y responsabilidades entre distintos órganos y niveles jurisdiccionales. No sorprende que el interés se centre a menudo en la eficacia de diferentes formas de descentralización educativa para mejorar los resultados del aprendizaje de los alumnos.

En este contexto, uno de los hallazgos más significativos es que los mecanismos claros de rendición de cuentas se vuelven mucho más importantes con una mayor descentralización, una transferencia de autoridad y una atribución real a las escuelas de autonomía y una facultad de decisión. La revisión de las acciones de reformas en distintos países muestra que la descentralización, la transferencia de autoridad y una mayor autonomía (*i.e.* para las escuelas) no generan necesariamente un mejor logro de los alumnos, a menos de que haya incentivos para las autoridades locales que van a rendir cuentas (The World Bank, 2004a, 2004b). Por ende, los incentivos se consideran como una forma de poner en práctica un marco más amplio de rendición de cuentas. A partir de las investigaciones y prácticas internacionales, es evidente que en donde hay un marco bien establecido de rendición de cuentas, la transferencia de autoridad de los procesos educativos favorece más el aumento del logro de los alumnos y es menos perjudicial para la equidad de oportunidades educativas de todos los alumnos, incluso aquellos de antecedentes socioeconómicos desfavorables.

Los acuerdos específicos para los incentivos docentes entre y al interior de distintos niveles de autoridad, adoptan múltiples formas. Por ejemplo, en los Estados Unidos de América, de cuatro de los sistemas de incentivos más estudiados y conocidos, dos son el resultado de acuerdos entre el estado y múltiples distritos y la sección local del sindicato magisterial (Dénver ProComp), y uno es un ejemplo de sistema con fondos federales⁴ en un solo distrito (Chicago). Por ende, el reto para los países es identificar y diseñar el acuerdo de transferencia de responsabilidades más eficaz y adecuado entre las jurisdicciones y órganos pertinentes. Esto cobra una importancia particular para el establecimiento de estándares de contenido y desempeño. Aunque las buenas prácticas de enseñanza pueden variar dependiendo de las condiciones locales, es mejor que algunos elementos, como los estándares de contenido y las reformas curriculares, sean responsabilidad del órgano central encargado de tomar decisiones. Esto garantizará que haya una visión unificada de lo que se espera sepan los alumnos y cómo hacerlo. De igual modo, los estándares de los maestros pueden servir como un marco de evaluación docente justo y válido si son establecidos con claridad en el ámbito central o nacional, siempre y cuando se contemplen los aportes y consideraciones de una amplia variedad de maestros y contextos de enseñanza. Así, para sistemas federales y descentralizados, las reformas de políticas educativas deben abordar necesariamente el tema de la transferencia de autoridad formal y *de facto*. Esto implica consideraciones que trascienden la relación de una simple rendición de cuentas bilateral basada en el financiamiento (*i.e.* fondos federales e implementación estatal/local). Quizá lo más importante sea que los sistemas educativos deben hallar el equilibrio correcto entre estándares y directrices establecidas nacional y centralmente, y la aplicación e innovación locales.

7.3 PROGRAMAS PILOTO, MONITOREO Y EVALUACIÓN DE INCENTIVOS

Además del análisis continuo sobre el diseño, la eficacia y las consecuencias de los esquemas de incentivos para los maestros, las prácticas internacionales destacan con claridad el valor de poner a prueba, evaluar y supervisar cuidadosamente tales iniciativas de reforma antes y durante una implementación total. Con base en las prácticas internacionales sobre el diseño, la planeación, la implementación y la evaluación de los programas de incentivos, el Apéndice al final de este capítulo ofrece ejemplos de las mejores prácticas y recomendaciones específicas sobre las pruebas piloto, el monitoreo y la evaluación de estas iniciativas de política pública. Los siguientes elementos destacan la importancia de efectuar programas piloto antes de implementar por completo programas nacionales, y se encuentran resumidos en el cuadro 7.1.

- **Los ejercicios de programas piloto diseñados, supervisados y evaluados cuidadosamente proporcionan evidencias objetivas sobre las formas más eficaces y eficientes de implementar esquemas de incentivos docentes.** Por ejemplo, los programas piloto de incentivos realizados en India y en Israel⁵ muestran que los ensayos evaluados con rigor brindan evidencias sobre la eficacia de los programas de incentivos docentes. Así, una evidencia objetiva y confiable ofrece información para las discusiones y compromisos entre los actores implicados sobre los elementos específicos del programa. Por ejemplo, cuando no se haya podido llegar a un acuerdo con los sindicatos magisteriales sobre ciertos elementos, estos se pueden poner a prueba empíricamente en condiciones bien diseñadas en las jurisdicciones participantes.

- **Las experiencias de los programas piloto identifican y abordan resultados potencialmente indeseables de los esquemas de incentivos docentes antes de poner en marcha la aplicación completa de un programa nacional.** Las experiencias en India, Israel, Kenia, el Reino Unido y los Estados Unidos de América, muestran que algunos efectos perversos que se abordan de manera adecuada mediante los programas piloto son: el “enseñar para la prueba”; la exclusión de alumnos de bajo desempeño o en desventaja; la corrupción o falsificación de mediciones de desempeño; y el descuido de temas que no aparecen en las pruebas, como ciudadanía, música o artes.
- **Los programas piloto identifican y cuantifican gastos inesperados durante la implementación.** También identifican de manera más realista los gastos de un programa utilizando sólo una parte de lo que costaría implementar y evaluar un programa nacional. Los elementos del programa ineficaces o ineficientes se identifican en la etapa piloto y se corrigen, o descartan, en el diseño de un programa nacional más amplio, lo que lo hace más rentable.
- **Los ejercicios piloto efectuados en contextos específicos se diseñan para compensar las condiciones que tal vez no estén presentes en el ámbito nacional,** y también dan una idea sobre cómo extender poco a poco estas condiciones a mayor escala. Por ejemplo, la información actualizada y precisa sobre el personal docente y los grupos de alumnos de una escuela debe ser recopilada y validada con rigor en las escuelas participantes en las muestras piloto.

Cuadro 7.1

Resumen de los beneficios de realizar programas piloto

Oportunidad de evaluar la viabilidad de los elementos del programa (*i.e.* medidas de desempeño docente, criterios y tipo/monto de las recompensas).

Oportunidad de identificar la eficacia del programa para alcanzar los resultados deseados en el comportamiento de los maestros y en el aprendizaje de los alumnos.

La escala limitada permite un proceso de evaluación controlado y robusto.

Oportunidad de poner a prueba prácticas innovadoras, si es que los programas pilotos por evaluar están bien diseñados.

Oportunidad de identificar y evaluar efectos negativos no deseados de los elementos del programa.

Oportunidad de evaluar el costo real de los elementos del programa, así como los gastos de implementación imprevistos.

Oportunidad de identificar el control de calidad de los datos necesario para una implementación eficaz del programa.

Oportunidad de implementar distintos modelos piloto simultáneos y en distintos contextos para comparar y contrastar su eficacia.

Oportunidad de identificar y ajustar el grado de apoyo o resistencia de los actores implicados (*i.e.* maestros, directores, administradores y secciones locales de sindicatos).

Ofrecer evidencias y datos para apoyar programas de mayor escala (es decir, estatales o nacionales).

Fuentes: Elaborado y adaptado a partir de H. Rogers, 2009; OCDE, 2009a, y con base en discusiones del Consejo Asesor de la OCDE sobre Políticas de Evaluación e Incentivos Docentes.

7.4 CONSIDERACIONES PARA MÉXICO

Si bien las recompensas al desempeño se han empleado con eficacia en otras áreas laborales, su uso reciente en el sector educativo, en particular para los maestros, aún está en etapa de exploración, supervisión y evaluación. De tal forma, la SEP, las autoridades educativas estatales y los actores implicados necesitarán determinar cuál es la combinación específica de incentivos y estímulos monetarios y no monetarios más eficaces para México. Sin embargo, independientemente de las recompensas o consecuencias que se vinculen a los resultados, para que los maestros puedan ser considerados como *eficaces*, sus alumnos deben demostrar niveles satisfactorios de

crecimiento; al mismo tiempo, ningún maestro debe ser considerado *ineficaz* si sus alumnos muestran niveles satisfactorios de crecimiento. Con base en la revisión de las principales características y ejemplos de incentivos y estímulos docentes de diversos países, a continuación se presenta un resumen de las recomendaciones para que México establezca políticas eficaces de incentivos y estímulos para maestros en servicio.

- Para conseguir una política eficaz y sustentable de incentivos para maestros en servicio, los siguientes cinco principios deben marcar la pauta de su elaboración:
 1. Los incentivos deben reflejar la calidad de la enseñanza. El criterio para decir que un programa de incentivos tiene éxito no sería tanto pagar mejor a los maestros que se desempeñan mejor, sino la contribución de los maestros para mejorar los resultados del aprendizaje de los alumnos.
 2. El sistema de incentivos debe reconocer y apoyar al maestro individual, al equipo de maestros en una escuela y a la profesión en su conjunto. Los incentivos deben integrarse a un sistema que apoye la mejora continua tanto de alumnos, maestros y escuelas, como del sistema educativo. En el largo plazo, los incentivos basados en pruebas deben complementarse con una sólida capacidad de manejo de recursos humanos en escuelas y ámbitos locales que evalúe con precisión la calidad del trabajo, utilizando métodos fuertes de validación y corroboración externas llevadas a cabo por el gobierno y la profesión docente.
 3. El sistema de incentivos debe basarse en una comprensión sólida de lo que motiva a los maestros, y abarcar múltiples dimensiones de motivación, con el propósito de fomentar un ambiente laboral atractivo, crear y facilitar el avance en la trayectoria profesional, ofrecer acceso al desarrollo profesional, e identificar y promover prácticas eficaces de enseñanza. Por ende, los incentivos y estímulos deben considerar incentivos financieros y no pecuniarios, como condiciones laborales, material suficiente para escuelas y salones de clase, reconocimiento social, mejores oportunidades de formación y desarrollo profesional, o una combinación de lo anterior.
 4. El sistema de incentivos debe proporcionar buenos mecanismos de retroalimentación y acceso al desarrollo profesional para garantizar así que los maestros que no reciban los incentivos comprendan qué pueden hacer para mejorar su desempeño y obtener incentivos para cambiar su conducta. También debe fomentar una cultura basada en evidencias y datos.
 5. El sistema de incentivos debe recompensar tanto el buen desempeño como la mejora relativa, y debe considerar el valor agregado de maestros y escuelas sin tomar en cuenta los factores socioeconómicos. Sin embargo, aunque se estén elaborando modelos analíticos de valor agregado, se pueden emplear métodos más sencillos para constatar que los alumnos, escuelas y maestros están siendo comparados con contextos semejantes (*i.e.* modelos de estratificación socioeconómica y/o de logro contextualizado).
- ***Es importante distinguir con claridad las políticas públicas de incentivos para maestros en servicio de otros programas relacionados con la enseñanza que sean semejantes sin que ofrezcan de manera fundamental incentivos a los maestros para mejorar su desempeño.*** Las políticas de incentivos deben ser comunicadas a los profesionales de la enseñanza de manera clara antes de informarles sobre las evaluaciones y medidas que se emplearán para los premios. Además, todos los maestros elegibles deben tener una probabilidad de recibir recompensas por un desempeño sobresaliente que sea superior a cero, lo cual no es el caso actualmente en programas semejantes en México. Por tal motivo, será de particular importancia hallar un equilibrio entre las directrices nacionales para los incentivos (y al menos un financiamiento parcial), y la flexibilidad y la coparticipación a nivel estatal por parte de las autoridades educativas locales en los recursos (financieros o de otro tipo) para incentivos y estímulos para los maestros. Por último, es bastante recomendable realizar una prueba piloto de los posibles programas de incentivos para garantizar la viabilidad y rentabilidad del diseño de las políticas. Las pruebas piloto deben ser supervisadas y evaluadas con rigor para que sean útiles y valiosas, con líneas base, y bajo tanto control como sea posible (*i.e.* ensayos aleatorios o cuasi-experimentales, si las condiciones lo permiten).

- **Los incentivos para maestros en servicio en México deben motivar a cada maestro a que mejore su desempeño, sin perder de vista a la escuela como la unidad básica para la rendición de cuentas**, tomando en cuenta el estado actual y las perspectivas de los sistemas de datos, así como la calidad de la información disponible. Sin embargo, incluso con sistemas de datos e información que no son lo suficientemente robustos, las autoridades educativas locales pueden elaborar medidas para confirmar y validar la elegibilidad de los maestros para los incentivos (i.e. con inspecciones de campo y la validación de datos sobre la información de los alumnos, escuelas y maestros). Conforme se vaya elaborando un sistema de evaluación docente sólido y creíble, las políticas de incentivos pueden modificarse para garantizar que los maestros estén en posibilidad de recibir incentivos de manera individual en el futuro.⁶ Para los incentivos escolares, las escuelas deben rendir cuentas de manera pública por los recursos adicionales que reciban. Si las escuelas pueden decidir la asignación de los recursos provenientes de los incentivos, debe haber mecanismos que aseguren la transparencia y la participación progresiva de los actores implicados pertinentes, como es el caso de los padres de familia y los consejos escolares locales.
- **Los incentivos y estímulos económicos y no económicos para los maestros deben basarse en un proceso de medición y evaluación justo y adecuado.** Debido a la gran diversidad del sistema educativo mexicano, es necesario elaborar un proceso de evaluación válido y confiable que identifique a los maestros elegibles para los incentivos. El éxito de los incentivos se vincula directamente a la credibilidad e imparcialidad del proceso de medición y evaluación en que se basen. Se deben emplear modelos que tomen en cuenta la diversidad socioeconómica de los alumnos mexicanos, así como otros factores que influyen en gran medida en el desempeño de los alumnos al comparar las escuelas y a sus maestros, como pueden ser el español como segundo idioma o el origen étnico. Las escuelas y programas de educación especial, así como las escuelas de nivel preescolar, pueden ser evaluados con base en mediciones adecuadas del desempeño docente y aprendizaje del alumno, cuando sea posible. Dada la diversidad que existe entre y al interior de los estados, las políticas públicas de incentivos también deben considerar un bono adicional relativo para escuelas rurales desfavorecidas, en contraparte de las escuelas urbanas o rurales sin desventajas. Asimismo, los incentivos deben apoyar la mejora continua de las escuelas y los maestros en todo el espectro del desempeño.

APÉNDICE 7A

LAS PRUEBAS PILOTO, EL MONITOREO Y LA EVALUACIÓN DE INCENTIVOS DOCENTES: PRÁCTICAS RECOMENDADAS

El propósito de este Apéndice es proporcionar a los responsables de formular políticas educativas, así como a su personal, un panorama de cuestiones relativas a las pruebas piloto, el monitoreo y la evaluación de los programas de incentivos docentes. Desde un principio se deben plantear las preguntas adecuadas sobre las metas del programa: ¿Cómo está siendo implementado? ¿Cómo responden los maestros a los elementos particulares del programa? ¿Qué consecuencias imprevistas se observan? Combinadas con cuestiones causales, que buscan información sobre los resultados del programa, estas preguntas permitirán una comprensión sólida en torno a la relación entre el propósito del programa de incentivos, lo que está logrando en la práctica y los resultados que genera. La realización de pilotos, el monitoreo y la evaluación son procesos distintos, aunque dependientes entre ellos, de un programa de incentivos bien planeado e implementado. Es importante considerar que los incentivos docentes son un área innovadora, y algunos de los ejemplos más pertinentes apenas comienzan a ser evaluados para saber qué es lo que funciona, en qué escenarios y en qué condiciones.

Recuadro 7A.1 Definiciones importantes

¿Qué significa realizar una **prueba piloto**? En esencia, es la puesta en marcha, controlada y sistemática, de un programa en un escenario restringido antes de que sea implementado por completo en todo un sistema educativo. Un programa piloto permite examinar la manera en que un programa se traslada del diseño a la implementación, y el impacto que tiene la implementación en los resultados tales como las actitudes y la conducta de los maestros, así como el aprendizaje de los alumnos. Mediante un programa piloto, es posible saber cómo difiere la realidad de lo que se previó, y cómo adaptar la implementación y el diseño para alcanzar mejor esas metas. Además, un sistema educativo puede identificar recursos que quizá se necesiten para apoyar una mejor implementación, como por ejemplo una formación para los educadores, un plan de comunicación o mayor capacidad técnica.

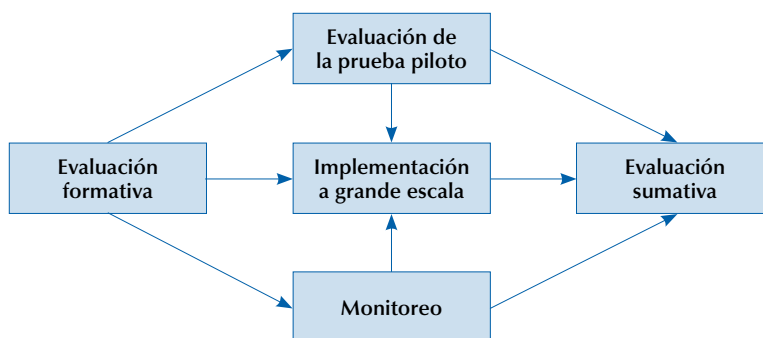
¿Qué significa **el monitoreo**? Mientras la evaluación aborda cuestiones empíricas relacionadas con la naturaleza, la calidad y los resultados de un programa, el monitoreo es un proceso que asegura que los programas se implementen de acuerdo con su diseño y ofrezcan garantías de calidad para su operación. Los sistemas de monitoreo buscan garantizar que el programa de incentivos funcione como fue previsto, e identificar las maneras que el funcionamiento sea distinto. El monitoreo puede incluir aspectos de operación tales como las campañas de comunicación y el pago de los premios dentro de los plazos estipulados. El monitoreo también sirve como proceso de control de calidad para asegurar, por ejemplo, que los sistemas de datos necesarios para la implementación del pago como incentivo funcionen de manera adecuada.

¿Qué significa **la evaluación**? La evaluación del programa no es algo que esté completamente separado del hecho de ponerla a prueba: todo programa piloto fructífero debe ir acompañado de una evaluación exhaustiva y rigurosa para aprender de la experiencia y realizar ajustes antes de lanzar el programa completo. El propósito de la evaluación – sea de un programa piloto o de la versión a grande escala – es averiguar empíricamente *qué funciona y por qué*, mediante métodos adecuados y justificables. Hay tres categorías de evaluación: las evaluaciones descriptivas se centran en describir el contexto, mientras que las evaluaciones de procesos ilustran la naturaleza de la implementación y de las experiencias del programa. Por otra parte, las evaluaciones causales identifican el impacto del programa. En última instancia, una evaluación bien diseñada explora la naturaleza de la implementación, el diseño del programa, el impacto en los alumnos, maestros y otros actores implicados, y bajo qué condiciones.

Las lecciones que se aprenden durante una prueba piloto influyen en la manera de implementar un programa, permiten conocer las áreas de implementación que requieran una supervisión rigurosa y ofrecen consideraciones para la evaluación completa, incluyendo áreas de interés, métodos y previsión de desafíos potenciales. Una vez implementado el programa por completo, el proceso de monitoreo, junto con las lecciones de la evaluación formativa, sirve para mejorar la implementación. Además, los resultados de la evaluación formativa pueden arrojar luz sobre áreas que necesiten una mayor o menor supervisión, mientras que los problemas identificados mediante el monitoreo también deben ser tomados en cuenta en la creación de métodos de investigación de evaluación y en la identificación de temas de interés. Todas estas acciones y procesos se encuentran al final en una evaluación sumativa que identifica los resultados del programa que se obtuvieron con base en su implementación (gráfica 7.1).

Gráfica 7.1

Relación entre la prueba piloto, el monitoreo y la evaluación de un programa



Fuente: Springer, 2010.

ENFOQUE Y PROPÓSITO DE LA EVALUACIÓN

Como se presenta en el cuadro 7A.1, hay tres categorías de preguntas que pueden abordarse en una evaluación: *descriptiva*, *de procesos* y *causal* (McEwan, 2008; Shavelson y Towne, 2002). El cuadro 7A.1 brinda algunos ejemplos sobre la forma de aplicar estas categorías de preguntas a la evaluación de un programa de incentivos, e incluye preguntas que suelen abordar los investigadores en este tema.

Cabe observar que estas categorías no deben considerarse de manera aislada. Por ejemplo, se puede formular un argumento causal sólido cuando una evaluación toma en cuenta el contexto y los procesos en que opera un programa. De hecho, hay quienes sostienen que el éxito de un programa puede depender tanto de la realidad de la implementación y apoyo del programa como del diseño del programa en sí mismo (Goldhaber, 2009).

DISEÑO Y MÉTODOS DE EVALUACIÓN

Panorama de métodos de evaluación

Para abordar las *preguntas descriptivas*, los evaluadores suelen emplear técnicas cuantitativas básicas, como son las estadísticas descriptivas, que miden la tendencia central de las variables, la dispersión y frecuencia de las variables, y la correlación (o asociación) de las variables entre sí (McEwan, 2008; Tukey, 1977). Por ejemplo, en el contexto de describir las características de las escuelas participantes en un programa de incentivos, estas técnicas descriptivas explican la dispersión de las escuelas por el nivel de pobreza, por el registro de desempeño anterior o –en promedio– por el tipo de alumno de las escuelas participantes. Mediante la correlación, es posible ver si las características de las escuelas se relacionan o no con la probabilidad de participar en el programa de incentivos.

Cuadro 7A.1

Enfoque y propósito de las preguntas de evaluación en el contexto de programas de incentivos

Tipo de pregunta	Preguntas clave
Descriptiva	<ul style="list-style-type: none"> – ¿Cuáles son las características de los centros que participan en el programa de pago como incentivo (<i>i.e.</i> situación de pobreza, registro de desempeño, tipo de alumno atendido)? – ¿Cuáles son las políticas públicas y el contexto político en los que se implementa el programa de pago como incentivo (<i>i.e.</i> iniciativas recientes de reformas, postura de los actores políticos claves respecto al pago como incentivo, cambios legislativos)? – ¿Cuáles son las directrices políticas que moldean las características de diseño del programa de pago como incentivo (<i>i.e.</i> contrato de jerarquización o de desempeño, monto del bono, mediciones del desempeño, premios individuales o grupales)? – ¿Cuál es la naturaleza del financiamiento para el programa de pago como incentivo (<i>i.e.</i> fuentes de financiamiento, montos, sustentabilidad)?
De proceso	<ul style="list-style-type: none"> – ¿Cuáles son las experiencias de los participantes durante el proceso de implementación (<i>i.e.</i> cómo se diseñó el programa, qué funcionó de acuerdo con el plan, qué problemas imprevistos surgieron)? – ¿Qué cambios se hicieron a la infraestructura o apoyo/recursos del sistema para implementar el programa de pago como incentivo (<i>i.e.</i> modificaciones a los sistemas de datos, prácticas de recursos humanos, apoyo técnico y formación)? – ¿Cuáles son las características de la implementación del programa; las entienden los participantes (<i>i.e.</i> cómo se mide el desempeño, cómo se determina el monto de los premios, cuándo se espera el pago)?
Causal	<ul style="list-style-type: none"> – ¿Cuál es el impacto del programa de incentivos en las actitudes y conducta de los maestros (<i>i.e.</i> percepciones sobre la imparcialidad y utilidad del programa, influencia en la práctica de la instrucción, apoyo a y retención de los maestros)? – ¿Cuál es el impacto del programa en la dinámica organizacional al interior de las escuelas (<i>i.e.</i> colaboración o competencia entre educadores, prácticas de liderazgo y toma de decisiones)? – ¿Cuál es el impacto del programa en los resultados de aprendizaje de los alumnos (<i>i.e.</i> ganancias en el logro de los alumnos, tipos de alumnos que experimentan un mayor impacto, materias que experimentan un mayor impacto)?

Fuente: Springer, 2010.

Las *preguntas de procesos* suelen abordarse mediante una combinación de métodos, sobre todo cualitativos (McEwan, 2008; Patton, 2002). Los estudios de caso, las entrevistas y las visitas/observaciones en los centros de trabajo son invaluable para reunir detalles sobre la forma en que se elabora e implementa un plan de pago como incentivo para docentes y que tal vez sean difíciles de recopilar únicamente a partir de fuentes de datos cuantitativos (*i.e.* registros administrativos, resultados de estudios). Sin embargo, una imagen más completa y validada de implementación se beneficia sin duda de la triangulación de numerosas fuentes de datos, tanto cuantitativos como cualitativos. A modo de ejemplo, al examinar las características de implementación de un programa, los evaluadores pueden realizar entrevistas al equipo de diseño o de supervisión de un programa para conocer el proceso de diseño del programa y las fases de implementación. Las entrevistas con los maestros pueden indicar si la percepción que tienen refleja el propósito real del programa, mientras que una revisión de los registros administrativos puede revelar la manera en que realmente se aplican los componentes de los incentivos (*i.e.* los resultados de la evaluación del educador o el pago de premios).

Por lo general, las *preguntas causales* reciben la mayor atención porque pretenden determinar si un programa funciona o no. Para los responsables de formular políticas públicas y otros profesionales, esto reviste la mayor importancia al momento de determinar el tiempo de vida de un programa, en particular en el contexto actual de las políticas educativas impulsadas por la rendición de cuentas y los resultados. En un contexto de pago como incentivo, se puede plantear la pertinencia de la investigación causal si la participación de una escuela en un programa de pago como incentivo da por resultado una mayor retención de maestros eficientes que el de una escuela comparable no participante. La meta final de la investigación causal en este ejemplo es identificar la diferencia de retención de maestros eficientes entre la escuela que participa en la intervención y la misma escuela si no hubiese participado en un programa de pago como incentivo.

Los métodos de la investigación causal son casi siempre de carácter cuantitativo y se agrupan en una de las tres siguientes categorías: no experimentales, cuasi-experimentales y experimentales. McEwan (2008) resumió de forma sucinta la diferencia fundamental entre estos métodos: “La diferencia esencial entre estas categorías es el grado de control que ejerce el investigador sobre quien es asignado al programa/tratamiento y quien es asignado al grupo control/de comparación”. El cuadro 7A.2 presenta un resumen de las categorías de investigación causal, de mayor a menor calidad.

Cuadro 7A.2

Distintos diseños de evaluación para investigar el impacto del programa y las intervenciones de las políticas educativas

Categoría	Grado	Descripción
Diseño experimental	Mayor calidad	Asignación aleatoria a la condición de control y de tratamiento (<i>i.e.</i> como cuando el investigador decide al lanzar al aire una moneda).
Cuasi-experimental	Calidad moderada	Uso de controles de concordancia y estadísticos, o una estrategia semejante, para establecer el tratamiento y las condiciones de comparación en ausencia de la aleatoriedad.
No experimental	Menor calidad	Estudio de correlación u observación, sin asignación aleatoria o condición de comparación. El investigador no influye en la asignación, que se debe por completo a la autoselección.

Fuente: Springer, 2010.

Métodos estadísticos

Dos controles estadísticos comunes son el análisis de regresión y el ajuste (comparación) de resultados tendenciales (*propensity score matching* en inglés). Con el análisis de regresión se controlan estadísticamente los factores que pueden explicar las diferencias en los resultados (que no se reduzcan sólo al tratamiento). Por ejemplo, si se trata de aislar en el logro de los alumnos el efecto de un programa de pago como incentivo basado en la escuela, el análisis de regresión puede integrar variables como la pobreza o el desempeño anterior del alumno, y la experiencia o el desgaste del maestro en el ámbito escolar. No obstante, sólo son observables los controles usados en la regresión.

El ajuste (comparación) de resultados tendenciales implica que el evaluador debe elaborar una puntuación que represente la probabilidad prevista de que una unidad reciba un tratamiento. Después, basándose únicamente en las puntuaciones elaboradas de la probabilidad, se hace concordar cada unidad en el grupo de tratamiento con una unidad de un grupo de comparación. Si no hay concordancia para una unidad en el grupo de tratamiento, esa unidad particular se retira del análisis, de modo que se reduzcan las diferencias entre el grupo de tratamiento y el de comparación. Luego, el evaluador estima el efecto de un programa o política pública al comparar la diferencia promedio de resultados entre las unidades concordadas en el grupo de tratamiento y en el de comparación (McEwan, 2008). Para hacer que la puntuación de la propensión concuerde en el contexto de un programa de pago como incentivo, hay que imaginar que un evaluador elabora una puntuación (con base en variables observadas) que representa la probabilidad de que una escuela participe en el programa. A continuación, un evaluador hallaría, por cada escuela en el programa, una escuela semejante no en el programa, pero con un valor de puntuación semejante. En este punto, un análisis de las diferencias de los índices de asistencia de los maestros entre las escuelas concordadas, por citar un ejemplo, sería el resultado del programa de pago como incentivo. Sin embargo, como sucede con el análisis de regresión, la comparación de resultados tendenciales no garantiza que se tomen en cuenta todas las variables no observadas.

Requisitos de una evaluación

Si bien hay pocos argumentos en contra del mérito de un programa riguroso de evaluación que proporcione una retroalimentación útil sobre el funcionamiento de un programa, a menudo los escenarios y condiciones, así como las necesidades prácticas y las condiciones previas para efectuar una evaluación, provocan que se

realice de manera deficiente o no se lleve a cabo completamente. Uno de los principales argumentos es que un sistema educativo debe comprometerse con el aprendizaje continuo apoyado en evidencias sólidas. En estos sistemas, los actores implicados saludan las acciones sistemáticas, rigurosas y transparentes que buscan entender qué es lo que funciona, y por qué, con el fin de hacer los cambios necesarios a través del tiempo. Los actores implicados también comprenderán que toma tiempo aprender cuál es la eficacia del programa. A corto plazo, una evaluación sería capaz de identificar cambios en la práctica o actitudes profesionales dentro de una escuela, aunque pueda tomar más tiempo determinar los resultados a largo plazo, como son la calidad de una planta laboral docente o el logro de los alumnos.

Programas piloto y experimentos de incentivos docentes: Ejemplos internacionales

Una de las estrategias más comunes es implementar y evaluar un programa piloto antes de hacer una implementación completa. Los siguientes son algunos ejemplos internacionales:

- El experimento del Proyecto de Incentivos de Enseñanza (*Project on Incentives in Teaching*, POINT por sus siglas en inglés) fue un estudio de una duración de tres años sobre los maestros de matemáticas de nivel medio, sus alumnos y escuelas, efectuado por el Centro Nacional de Incentivos por Desempeño (*National Center on Performance Incentives*, NCPI por sus siglas en inglés), realizado con fondos federales. El experimento POINT se llevó a cabo en un gran distrito urbano de escuelas públicas en el estado de Tennessee, Estados Unidos de América (EUA), con el propósito de estudiar qué efectos podría tener en los resultados de alumnos pagar a los maestros bonos por incentivos de hasta 15 000 dólares estadounidenses al año, tomando como base el aumento de las puntuaciones en las pruebas de los alumnos. Los maestros voluntarios fueron asignados durante el estudio a la condición de tratamiento o a la de control de manera aleatoria. Los bonos de los maestros en la condición de tratamiento se basaron en dos criterios: el avance en un año de los alumnos de un maestro de matemáticas, medido por el aumento de las puntuaciones de sus pruebas; y el avance en un año de los alumnos de un maestro que no fuera de matemáticas, medido por los aumentos en las puntuaciones de sus pruebas. Los resultados de este experimento aún están en proceso de elaboración (Springer, 2010).
- El NCPI se encuentra actualmente en el segundo año de un experimento sobre modelos de incentivos, aplicados a nivel de equipo de maestros, en el Distrito de Escuelas Independientes de Round Rock (*Round Rock Independent School District*, RRISD por sus siglas en inglés), en Texas, EUA. El distrito dispone a sus maestros de educación media en equipos interdisciplinarios por grado, los cuales supervisan las experiencias de aprendizaje de distintos grupos de alumnos. Este estudio permite explorar la forma en que los premios a equipos impactan el logro de los alumnos en áreas temáticas básicas (es decir, matemáticas, lectura, escritura, estudios sociales y ciencias). También permite examinar el impacto de los premios a equipos en la conducta de los maestros, en la dinámica institucional y organizacional, y en muchos otros aspectos de los procesos escolares. El estudio es en particular pertinente en las discusiones actuales sobre políticas, ya que muchos individuos y grupos aseguran que un equipo de grado-año es el nivel más adecuado para recompensar el desempeño en el sector educativo.

Durante el año lectivo 2007-2008, la Ciudad de Nueva York (NYC por sus siglas en inglés) y la Federación Unida de Maestros (*United Federation of Teachers*, UFT por sus siglas en inglés) diseñaron el primer programa de pago como incentivo para escuelas de la ciudad. Se recaudaron unos 20 millones de dólares estadounidenses de fuentes privadas para apoyar la iniciativa piloto. En noviembre de 2007 se seleccionaron de manera aleatoria 240 escuelas públicas de NYC (15% del total) para participar en un conjunto de escuelas con grandes necesidades, definidas por la clasificación de competencia promedio en áreas temáticas básicas, índices de pobreza, datos demográficos de los alumnos y porcentajes de alumnos que aprenden el idioma inglés y de educación especial. De esas 240 escuelas, 205 (un 86%) accedieron a participar. Al comienzo del año lectivo 2008-2009, el programa recibió fondos públicos y se amplió para incluir a más de 400 escuelas (30% de todas las escuelas públicas de NYC).

Las escuelas elegibles optaron por participar en el programa mediante el voto de un comité de compensación escolar emitido durante el año lectivo 2007-2008. Cada escuela diseñó objetivos de reporte de progreso para determinar su elegibilidad para los premios de incentivos escolares, distribuidos al final del año lectivo. Las escuelas que cumplieron con todos los objetivos de desempeño pudieron obtener fondos suficientes para que todos los empleados de tiempo completo pertenecientes a la UFT recibieran 3 000 dólares estadounidenses. Las escuelas que alcanzaron 75% de los objetivos obtuvieron fondos suficientes para que cada empleado reciba 1 500 dólares estadounidenses. El comité de compensación de cada escuela decidió cómo distribuir los premios de incentivos entre los empleados.

Una evaluación de este programa piloto se encuentra en sus primeras fases, y se tienen resultados disponibles del primer año de implementación del programa. Springer y Winters (2009) observaron que tras un año de su implementación, el programa de incentivos para escuelas tuvo poco impacto en el desempeño de los alumnos o en el ambiente interno de las escuelas. Sin embargo, los investigadores señalan que es difícil determinar la eficacia de un programa cuya vigencia es apenas de un año. Los resultados de los años subsecuentes del programa de NYC estarán listos próximamente.

Junto con la compañía de Investigación Matemática en Políticas Públicas (*Mathematica Policy Research, Inc.*), los evaluadores diseñaron una evaluación experimental del Programa de Avance Docente (*Teacher Advancement Program, TAP* por sus siglas en inglés) en un subconjunto de escuelas públicas del sistema de Escuelas Públicas de Chicago (*Chicago Public Schools, CPS* por sus siglas en inglés), en los Estados Unidos de América. Modelo común en ese país, el TAP es un sistema amplio para mejorar la calidad docente que incluye un componente de pago como incentivo para los maestros. Los otros componentes de la calidad docente del TAP incluyen un escalafón profesional y desarrollo profesional continuo impulsado por datos. Con el modelo del TAP, los maestros obtienen pagos y responsabilidades adicionales al ascender al grado de mentor o maestro experto, y ganan bonos de incentivos anuales con base en una combinación de su valor agregado con el logro del alumno y el desempeño observado en el salón de clases. El modelo también incluye reuniones semanales con grupos de maestros y evaluaciones periódicas del salón de clases por parte de un equipo de liderazgo escolar para ayudar a que los maestros alcancen sus metas de desempeño. El TAP se implementó en más de 200 escuelas en todo el país (Springer, 2010).

De las 16 escuelas primarias del CPS que solicitaron participar voluntariamente en el TAP de Chicago y lograron completar el proceso de selección, ocho fueron asignadas de manera aleatoria a un grupo de tratamiento y las otras ocho a un grupo control que retrasó la implementación hasta el siguiente año lectivo (2008-2009). Con el fin de complementar el análisis experimental, los investigadores emplearon procedimientos de concordancia de la propensión de puntuación para crear una muestra de comparación entre las más de 200 escuelas que se correspondían, de acuerdo con características previas a la intervención tales como el tamaño de la escuela, los datos escolares demográficos, el logro de los alumnos y la retención de los maestros (Glazerman, McKie y Carey, 2009; Glazerman y Seifullah, 2010).

En el reporte del primer año del experimento, Glazerman *et al.* (2009) observaron que la implementación del TAP generó algunos cambios en la dinámica organizacional al interior de las escuelas, a saber, un mayor apoyo de los mentores hacia los maestros. No hubo un impacto detectable en las puntuaciones de logro de los alumnos, pero sí una mayor retención de maestros en las escuelas con el TAP, aunque no en todo el distrito. Tras el segundo año, los investigadores aún no detectaron diferencia alguna en el aumento del logro de los alumnos entre las escuelas tratadas y las del grupo control. Tampoco hubo un impacto significativo en la retención de maestros en las escuelas con el TAP o en el distrito (Glazerman y Seifullah, 2010). La investigación se mantendrá para examinar los resultados en los próximos años.

En asociación con la compañía de Investigación Matemática en Políticas Públicas (*Mathematica Policy Research, Inc.*), y con fondos del Departamento de Educación de los Estados Unidos de América, se diseñó un experimento aleatorio llamado Iniciativa de Transferencia de Talento (*Talent Transfer Initiative, TTI* por sus

siglas en inglés) para poner a prueba una política de pago como incentivo que ofrece un pago adicional a los maestros más eficientes si aceptan ser asignados en escuelas de bajo desempeño.⁷ En siete distritos escolares grandes y diversos en todo el país, se aplicaron análisis de valor agregado a los datos de tres años de crecimiento en el logro de los alumnos para identificar al mejor 20% de maestros de inglés de primaria y secundaria, y de matemáticas de secundaria (Springer, 2010).

Quienes implementaron la TTI ofrecieron 20 000 dólares estadounidenses a los maestros para ser asignados en escuelas de bajos logros y facilitaron el proceso. Al mismo tiempo, el programa identificó un grupo de escuelas de bajo rendimiento con puestos docentes vacantes en los grados y materias buscados; de forma aleatoria, el equipo de investigadores asignó la mitad de ellos a un grupo de tratamiento y la otra mitad a un grupo control. Las escuelas en tratamiento fueron elegibles para contratar a un maestro de la TTI de excelente desempeño, a quien le propusieron 20 000 dólares estadounidenses por su asignación. Las escuelas del grupo control tuvieron que ocupar sus puestos vacantes mediante procedimientos normales. Si bien aún está en su fase temprana, el estudio seguirá durante dos años a los maestros transferidos y a los correspondientes maestros del grupo control.

En India, una serie de experimentos identificó resultados positivos en las políticas de pago como incentivo, en particular en el mejoramiento de la asistencia de los maestros y el desempeño de los alumnos en los exámenes. Mediante asignación aleatoria, Duflo, Hanna y Ryan (2007) estudiaron el efecto del pago como incentivo en instalaciones educativas rurales no oficiales, en donde la asistencia de los maestros era tradicionalmente muy baja. Se observó que el pago como incentivo redujo las ausencias de los maestros hasta en un 19%, y al final, que también mejoraron las puntuaciones en las pruebas y los índices de graduación de los alumnos.

Muralidharan y Sundararaman (2008) efectuaron otro experimento aleatorio en India. En este experimento de incentivos en escuelas primarias rurales gubernamentales, se designó a cuatro grupos experimentales. Los dos primeros grupos recibieron apoyos adicionales, uno en forma de otro docente asistente y el otro en forma de una beca en efectivo para adquirir más recursos escolares. El tercero y el cuarto grupos experimentales recibieron bonos de pago como incentivo al obtener mejores puntuaciones en las pruebas de los alumnos, uno con base en el desempeño de un grupo de maestros, el otro en el desempeño individual de los maestros. Todos los tratamientos (es decir, el docente asistente, la beca en efectivo y los bonos de pago como incentivo) fueron de igual valor para las escuelas. Muralidharan y Sundararaman (2008) observaron que, en comparación con las escuelas control, el pago como incentivo tuvo el mayor impacto en el logro de los alumnos, pero también que el logro de los alumnos fue aun mayor en las escuelas que recibieron apoyos adicionales en comparación con las escuelas control. En general, se apreció que los apoyos y el pago como incentivo eran ambos más rentables en las escuelas control que el *statu quo*, siendo el pago como incentivo el tratamiento más eficaz.

En Israel se efectuaron dos estudios de “torneos” de desempeño (Lavy, 2002; 2004).⁸ En ambos estudios, el programa se diseñó para elevar los índices de aprobación en exámenes finales de preparatoria en las escuelas de bajos recursos socioeconómicos. Aunque las escuelas no se asignaron de forma aleatoria a la condición de control o a la de tratamiento, ambos programas se implementaron conforme a tres reglas oficiales de asignación (grado, desempeño anterior e índice de inscripciones), lo que permitió una evaluación más rigurosa.⁹ El Experimento Israelí de Incentivos Docentes (*Israeli Teacher-Incentive Experiment*) también se diseñó cuidadosamente para reducir las conductas deshonestas u oportunistas por parte de los maestros y administradores escolares: las medidas de desempeño se basaron en el tamaño de la generación que se graduaba con el fin de desalentar a las escuelas a que fomentaran transferencias o exclusiones de alumnos desfavorecidos, colocándolos en registros ajenos a la matrícula.

El primer estudio de Lavy (2002) examinó el torneo en el que un grupo seleccionado de preparatorias de bajo desempeño compitió con base en el desempeño escolar. El tercio superior de escuelas, determinado por su mejora anual en puntuaciones de pruebas, recibió premios de 13 250 a 105 000 dólares estadounidenses. Los bonos para maestros fueron de 250 a 1 000 dólares, y se distribuyeron por igual a todos los maestros en

las escuelas “ganadoras”. Lavy observó un efecto positivo en las escuelas participantes en relación con el grupo de comparación no participante de escuelas de bajo desempeño. También concluyó que otorgar más recursos a las escuelas (25% de los premios para las escuelas fue para mejorar la infraestructura) contribuyó a un mejor desempeño de los alumnos.

El segundo estudio examinó el programa de bonos a maestros en lo individual, también efectuado como torneo (Lavy, 2004). En esencia, se clasificó a los maestros participantes con base en las contribuciones de valor agregado al logro de los alumnos en diversos exámenes finales; aquellos que se desempeñaron mejor recibieron los bonos. El programa abarcó a 629 maestros, de los cuales 302 recibieron premios. Los bonos fueron sustanciales: hasta 7 500 dólares estadounidenses por clase con respecto a un promedio salarial de 25 000 dólares. Los resultados indican un efecto positivo en el desempeño de los maestros participantes (es decir, tanto quienes recibieron bonos como quienes no lo hicieron), el cual ascendió en relación con un grupo de comparación de maestros que no participaron en el programa de incentivos.

En Kenia, Glewwe, Ilias y Kremer (2008) pusieron a prueba un programa de pago como incentivo basándose en grupos en ocho escuelas primarias seleccionadas al azar. El pago como incentivo se otorgó a las escuelas mejor clasificadas en las puntuaciones, o a las que mejoraron más. Los autores observaron que las puntuaciones en las pruebas de los alumnos mejoraron en el transcurso del programa de pago como incentivo, pero también que esas mejoras disminuyeron al término del programa (es decir, ya no hubo una diferencia detectable entre las puntuaciones de los alumnos en las escuelas en tratamiento y en las de control).

Diseño y consideraciones clave para la supervisión

Hasta ahora, este apéndice abordó el papel de los métodos para evaluar los programas y políticas de pago como incentivo. Otro punto importante que debe tomarse en cuenta es el de saber si un programa se implementa de acuerdo con su diseño. Si un sistema educativo no tiene los recursos o la capacidad para implementar un programa, o no lo implementa con fidelidad, se vuelve mucho más difícil que un evaluador sepa si el programa o las políticas son en verdad eficaces.

Esta sección presenta consideraciones clave para supervisar la capacidad de un sistema educativo para implementar un programa de pago como incentivo. Asimismo integra varias lecciones a partir de ejemplos recientes en varios países.

Los investigadores y los grupos encargados de formular políticas aprovecharon la oportunidad de explorar las experiencias de elaboración e implementación de dichas iniciativas de pago como incentivo. Las lecciones aprendidas –a partir de programas que tuvieron una implementación fructífera y otros que no– se agrupan en tres grandes categorías:

- *Políticas y liderazgo.* Esta categoría representa la medida en que las políticas públicas de un sistema educativo están alineadas y centradas en una visión común de la enseñanza y el aprendizaje. También aborda la necesidad de liderazgo, en los ámbitos de las escuelas y superiores, que no sea volátil y trabaje de manera concertada con un marco coherente de políticas.
- *Compromiso de los actores implicados.* Para diseñar e implementar bien un programa de pago como incentivo, un sistema educativo debe, desde el principio y durante toda la vida del programa, hacer partícipe a un conjunto amplio de actores implicados, incluyendo a los maestros. Este nivel de compromiso debe incluir comunicaciones frecuentes y transparentes, una participación en el diseño y la toma de decisiones, y oportunidades para que los actores implicados se retroalimenten y tengan una idea del modo en que viven esta experiencia.
- *Capacidad técnica.* Esta categoría representa la infraestructura y el apoyo técnicos que se necesitan materialmente para implementar y sostener un programa de pago como incentivo. Por ejemplo, un sistema educativo debe contar con sistemas de datos para medir con precisión el desempeño, junto con sistemas de nómina transparentes y exactos.

Recuadro 7A.2 Importancia de la participación de los actores educativos

Es primordial la participación temprana y continua de un amplio conjunto de los actores implicados en el diseño e implementación de un programa de incentivos. El compromiso significativo y continuo se garantiza mediante intercambios frecuentes y transparentes, una amplia participación de los actores implicados en las funciones del diseño y la toma de decisiones, y oportunidades para que los actores implicados se retroalimenten (Eckert, 2010; Lewis y Springer, 2009; OCDE, 2009). Entre los actores implicados se deben encontrar representantes de grupos de maestros, sindicatos magisteriales, encargados de escuelas y distritos, miembros de juntas escolares, personal de escuelas y distritos, y otros miembros de la comunidad (*i.e.* líderes de comercios, padres de familia). Al incluir estas múltiples perspectivas, es más probable que un programa de incentivos genere consensos en objetivos básicos y se acepte después de haber sido implementado (OCDE, 2009; Raue *et al.*, 2008).

El cuadro 7A.3 presenta un panorama de las principales cualidades ya descritas, así como ejemplos de preguntas y estrategias útiles para el monitoreo de estos controles de calidad antes y durante la implementación del programa.

Cuadro 7A.3

Enfoque y propósito de las preguntas de evaluación en el contexto de incentivos

Propósito	Preguntas para el monitoreo
Apoyo y consistencia del liderazgo	¿Los líderes educativos de la escuela, distrito y/o estado apoyan y están comprometidos con el programa de pago como incentivo? ¿Son «defensores» del programa?
Coherencia de las políticas	¿El sistema educativo tiene una estrategia amplia para mejorar la calidad docente? ¿El programa de pago como incentivo está alineado con esas metas? ¿El programa de pago como incentivo entra en conflicto o se opone a otras iniciativas de calidad docente?
Apoyo y sustentabilidad financieros	¿Hay un flujo constante de fondos para el programa de pago como incentivo? ¿Considera el sistema educativo cómo reasignar los fondos para abordar el pago como incentivo como componente de un esfuerzo más amplio de mejora de la calidad docente?
Comunicación con los actores implicados	¿Cuán a menudo y mediante qué mecanismos el sistema educativo se comunica con los actores implicados? ¿Cuál es el público al que se dirige, cuál es el contenido del material que se comunica? ¿Los actores implicados leen el material, lo comprenden y valoran?
Conciencia y comprensión del programa por los actores implicados	¿Los actores implicados –en particular los elegibles para recibir premios– saben de la existencia del programa? ¿Entienden cómo opera el programa y, en particular, cómo se evalúa su desempeño?
Participación de los actores implicados en el diseño y la implementación	¿Qué clase de actores implicados se encuentran en el comité de diseño; participan en alguna función activa? ¿Con qué mecanismos se hace partícipes activamente a los actores implicados en el diseño e implementación del programa de pago como incentivo? ¿Cómo consigue el sistema educativo una retroalimentación entre los actores implicados; qué se hace con la retroalimentación?
Datos que identifiquen quién está en el sistema educativo	¿El sistema de datos tiene un identificador único por alumno y por maestro? ¿El sistema de datos permite una concordancia entre el maestro y sus alumnos? ¿Hay un sistema de auditoría para verificar la calidad de estos datos?
Datos que identifiquen los resultados de los alumnos	¿El sistema de datos registra los resultados de los alumnos en lo individual, y es capaz de atribuir dichos resultados al maestro correspondiente? ¿Hay un sistema de auditoría para verificar la calidad de estos datos?
Sistema de pagos de los premios	¿Hay un sistema de pagos de los premios preciso y vinculado con los resultados de la evaluación del educador? ¿Los premios se pagan oportunamente y de manera que los educadores lo entiendan? ¿El sistema educativo tiene un procedimiento conciliatorio para los educadores que reclamen por el pago de su premio?

Fuente: Springer, 2010.

Recuadro 7A.3 La Campaña de Calidad de Datos

La Campaña de Calidad de Datos (*Data Quality Campaign*, DQC por sus siglas en inglés), esfuerzo colaborativo de varios socios en los Estados Unidos de América para mejorar los sistemas de datos estatales de la educación pública, emprendió una acción concertada para identificar las necesidades de datos y asesorar a los estados sobre las estrategias a seguir para obtener (los que la campaña considera) los “10 elementos esenciales” de un sistema de datos longitudinal estatal. Estos elementos se ordenan en tres categorías: *i*) identificar quién está en el sistema educativo, *ii*) entender los resultados de los alumnos y *iii*) supervisar la calidad del sistema de datos.¹⁰

¿Quién está en el sistema de educación pública? La DCQ recomienda que un sistema de datos longitudinal estatal incluya identificadores únicos para cada alumno y cada maestro. El identificador del alumno conecta los datos del alumno en varias bases de datos y captura la inscripción del alumno en escuelas y cursos, su información demográfica y su participación en programas como los de educación especial o de asistencia a la pobreza. Los identificadores únicos de los maestros deben relacionarlos con sus alumnos por clase y materia. El sistema de datos debe capturar información del maestro como es la escolaridad, experiencia y formación, y registrar su tránsito por escuelas/distritos.

¿Cuáles son los resultados de los alumnos? Todo sistema estatal de datos educativos también debe capturar información sobre los resultados de los alumnos, como son el desempeño en evaluaciones, la terminación de cursos, la graduación o el abandono, así como su trayectoria tras la educación secundaria. De manera más específica, la DQC recomienda que los resultados de las evaluaciones de los alumnos se recopilen a nivel del alumno y permitan, año con año, registrar las puntuaciones de las pruebas individuales para medir el crecimiento académico. Con el tiempo, se debe capturar igualmente la terminación de cursos y los grados a nivel individual del alumno, así como los resultados de los alumnos tras la educación secundaria (es decir, si continuaron su educación, y qué tipo de educación eligieron). Por último, un sistema de datos de buena calidad debe identificar del mismo modo a los alumnos que se inscribieron pero no se presentaron a las pruebas, de modo que con el tiempo se comprendan cuáles son los patrones y tipos de alumnos que no son evaluados.

¿Cuál es la calidad de los datos? La DQC señala que todo sistema de datos estatal debe acompañarse de un sistema de auditoría de datos que evalúe la calidad, la validez y la confiabilidad de los datos integrados al sistema longitudinal. Crear este sistema de auditoría implica establecer estándares de datos, una formación de personal escolar local a cargo de la provisión de datos, pruebas estadísticas para identificar errores, auditorías intermitentes en el lugar de registro de datos de los sistemas escolares locales, y una serie de consecuencias y/o recompensas para recalcar, frente a los sistemas escolares locales, la importancia de registrar e incorporar datos de buena calidad.

CONCLUSIÓN

Conforme aumenta la cantidad de países y autoridades educativas locales que consideran estrategias para mejorar el aprendizaje de los alumnos y la eficiencia de los educadores, las iniciativas de compensación e incentivos docentes asumen cada vez más un papel clave en las agendas de las reformas. La cuestión de implementar o no un pago como incentivo rápidamente pierde importancia frente a la cuestión sobre cómo implementar mejor un programa de incentivos para docentes. En el futuro será más importante que nunca que los estados y países que implementen programas de pago como incentivo incluyan programas piloto, evaluaciones y mecanismos de monitoreo y control de calidad rigurosos como componentes centrales de su labor.

Es difícil equilibrar las necesidades de la investigación, las políticas públicas y la práctica. Si bien una investigación y una evaluación rigurosas son vitales para mejorar el diseño y la implementación de los programas, las limitaciones de tiempo y recursos tal vez no permitan aplicar un programa piloto amplio antes de realizar una implementación a grande escala. Cuando no es viable una prueba piloto, es crítico que un sistema educativo ponga en marcha mecanismos para llevar a cabo una evaluación formativa continua, un monitoreo de calidad y una evaluación sumativa objetiva que permitan entender y modificar, según sea necesario, la naturaleza y las consecuencias de un programa de incentivos para docentes.

NOTAS

1. Los ejemplos, prácticas y recomendaciones sobre las políticas públicas relacionadas con la educación, selección y formación de los maestros escapa al alcance de este capítulo.
2. Los incentivos para los maestros en servicio basados en el desempeño son un enfoque innovador; los experimentos, ensayos y experiencias aún están en período de supervisión y evaluación. Sin embargo, las investigaciones son concluyentes acerca de los diferentes problemas con el esquema salarial único (OCDE, 2009a).
3. Este capítulo recurre en gran medida a la publicación hermana de este informe *Evaluación y reconocimiento de la calidad de los docentes: Prácticas internacionales* (OCDE, 2009), que resume las ideas y prácticas más comunes sobre los incentivos y estímulos para maestros, revisando los ejemplos de 23 países.
4. El Fondo de Incentivos Docentes (*Teacher Incentive Fund*), administrado por el Departamento de Educación estadounidense (USDOE) y dotado con un presupuesto de 400 millones de dólares en 2010, está destinado para dar fondos a los distritos con el fin de implementar sistemas de pago basados en mérito (USDOE, 2010).
5. En el Estudio de Evaluación Aleatorio Andhra Pradesh (*Andhra Pradesh Randomised Evaluation Study*, APREST) efectuado en India, Muralidharan y Sundararaman (2009) presentan conclusiones según las cuales los incentivos docentes mejoraron los resultados del aprendizaje de los alumnos, tanto en matemáticas como en lengua, tras sólo dos años de vigencia del programa. Lavy (2004) muestra que un programa piloto de un año en Israel mejoró el aprendizaje de los alumnos, y tuvo un impacto positivo en las conductas de los maestros, tales como los métodos de enseñanza, la enseñanza fuera de la escuela y una mejor respuesta a las necesidades de los alumnos.
6. Respecto a los montos adecuados de los incentivos, una revisión de los programas internacionales muestra que los incentivos docentes individuales varían de menos de un 1% a más de 360% del salario mensual (OCDE, 2009), aunque los expertos sugieren que para que los incentivos sean significativos, sin generar conductas no deseadas, estos deben representar entre 4% y 8% del salario anual.
7. En <http://talenttransferinitiative.org/index.html> hay más información sobre la Iniciativa de Transferencia de Talento.
8. Los premios de los torneos no se basan en un estándar absoluto, sino en un desempeño relativo.
9. Lavy empleó un diseño de discontinuidad de regresión en sus estudios en Israel sobre los efectos del pago como incentivo. Este diseño permite mediciones más precisas de los efectos de una intervención antes y después de que se implemente.
10. Hay más información sobre la Campaña de Calidad de Datos y los 10 elementos esenciales en www.dataqualitycampaign.org/survey/elements.

Referencias bibliográficas

Atkinson, A., S. Burgess, B. Croxson, P. Gregg, C. Propper, H. Slater y D. Wilson (2009), "Evaluating the Impact of Performance-Related Pay for Teachers in England", en *Labour Economics*, vol. 16, pp. 251-261.

Austin Independent School District (AISD) (2010), *AISD REACH Supporting and Rewarding Success in the Classroom Year 2: Evaluation Report II 2008-2009*, Austin, Texas.

Battelle for Kids (2009), *The Importance of Accurately Linking Instruction to Students to Determine Teacher Effectiveness*, Battelle for Kids.

Boruch, R., D. DeMoya y B. Snyder (2002), "The Importance of Randomized Field Trials in Education and Related Areas", en F. Mosteller y R. Boruch (eds.), *Evidence Matters: Randomized Trials in Education Research*, Brookings Institution Press, Washington, DC, pp. 50-79.

Buck, S. y J.P. Greene (2010), *Blocking, Diluting, and Co-opting Merit Pay*, documento de trabajo PEPG núms. 10-14, Programme on Education Policy and Governance, Cambridge, Massachusetts.

Buddin, R., D.F. McCaffrey, S.N. Kirby y N. Xia (2007), *Merit Pay for Florida Teachers: Design and Implementation Issues*, documento de trabajo WR-508-FEA, RAND Education.

Chamberlin, R., T. Wragg, G. Haynes y C. Wragg (2002), "Performance-Related Pay and the Teaching Profession: A Review of the Literature", en *Research Papers in Education*, vol. 17, núm. 1, pp. 31-49.

Coates, Y.D. (2009), *A Focused Analysis of Incentives Affecting Teacher Retention: What might Work and Why?*, American University, Washington, DC.

Community Training and Assistance Center (2004), *Catalyst for Change: Pay for Performance in Denver*, CTAC, Boston, Massachusetts.

Cornetto, K.M., L.S. Schmitt, K. Malerba y A. Herrera (2010), *AISD REACH Year 2 Evaluation Report II, 2008-2009*, Austin Independent School District, Austin, Texas.

Cresap, McCormick and Paget, Inc. (1984), *Teacher Incentives: A Tool for Effective Management*, Reston, National Association of Elementary School Principals, Washington, DC.

Darling-Hammond, L. (1997), *Doing What Matters Most: Investing in Quality Teaching*, National Commission on Teaching & America's Future, Nueva York.

Delaware Department of Education (2010), "Race to the Top: Application for Funding", CFDA núm. 84395A, Narrative, El Estado de Delaware.

Duflo, E., R. Hanna y S. Ryan (2007), *Monitoring Works: Getting Teachers to Come to School*, documento de trabajo 11880, National Bureau of Economic Research, Cambridge, Massachusetts.

Eckert, J. (2010), *Performance-Based Compensation: Design and Implementation at Six Teacher Incentive Fund Sites*, The Bill and Melinda Gates Foundation y The Joyce Foundation.

Ehrenberg, R. y R. Smith (1994), *Modern Labor Economics: Theory and Public Policy*, 5a ed., Harper Collins, Nueva York.

Ehrenberg, R.G. y G.T. Milkovich (1987), "Compensation and Firm Performance", en M. Kleiner et al. (eds.), *Human Resources and the Performance of Firms*, Industrial Relations Research Association, Madison, Wisconsin.

Gates Foundation (2010), *Working with Teachers to Develop Fair and Reliable Measures of Effective Teaching*, Gates Foundation, Seattle, Washington.

Glazerman, S., A. McKie y N. Carey (2009), *An Evaluation of the Teacher Advancement Program (TAP) in Chicago: Year One Impact Report*, Mathematica Policy Research, Inc.

- Glazerman, S. y A. Seifullah** (2010), *An Evaluation of the Teacher Advancement Program (TAP) in Chicago: Year Two Impact Report*, Mathematica Policy Research, Inc.
- Glewwe, P., N. Ilias y M. Kremer** (2008), *Teacher Incentives in Developing Countries: Recent Experimental Evidence from Kenya*, documento de trabajo, Harvard University, Cambridge, Massachusetts.
- Goldhaber, D.** (2009), "The Politics of Teacher Pay Reform", en M. G. Springer (ed.), *Performance Incentives: Their Growing Impact on American K-12 Education*, Brookings Institution Press, Washington, DC.
- Gonring, P., P. Teske y B. Jupp** (2007), *Pay-for-Performance Teacher Compensation: An Inside View of Denver's ProComp Plan*, Harvard Education Press, Cambridge, Massachusetts.
- Hanushek, E.A.** (1992), "The Trade-off between Child Quantity and Quality", en *Journal of Political Economy*, vol. 100, núm. 1, pp. 84-117.
- Heinrich, C. y G. Marschke** (2007), *Dynamics in Performance Measurement System Design and Implementation*, La Follette School of Public Affairs, University of Wisconsin, Madison, Wisconsin.
- Heyburn, S., J. Lewis y G. Ritter** (2010), *Compensation Reform and Design Preferences of Teacher Incentive Fund Grantees*, National Center on Performance Incentives, Nashville, Tennessee.
- HISD** (2009), *Houston Independent School District*, Michael & Susan Dell Foundation, Houston, Texas.
- Jacob, B. y M. Springer** (2007), *Teacher Attitudes on Pay for Performance: A Pilot Study*, documento de trabajo 2007-06, National Center on Performance Incentives.
- Johnson, J.** (2005), *State Financial Incentive Policies for Recruiting and Retaining Effective New Teachers in Hard-to-Staff Schools*, Education Commission of the States Clearinghouse on Teacher Incentives, Denver, Colorado.
- Koppich, J.** (2008), *Toward a More Comprehensive Model of Teacher Pay*, National Center on Performance Incentives, documento de trabajo 2008-6.
- Lavy, V.** (2002), "Evaluating the Effect of Teachers' Group Performance Incentives on Pupil Achievement", en *Journal of Political Economy*, vol. 110(6), pp. 1286-1317.
- Lavy, V.** (2004), *Performance Pay and Teachers' Effort, Productivity and Grading Ethics*, documento de trabajo NBER 10622, National Bureau of Economic Research.
- Lawler, E.E. III** (1981), *Pay and Organizational Effectiveness*, McGraw-Hill, Nueva York.
- Lazear, E.** (1996), "Performance Pay and Productivity", documento de trabajo NBER núm. W5672, National Bureau of Economic Research, Cambridge, Massachusetts, SSRN, <http://ssrn.com/abstract=225573>, consultado en abril de 2010.
- Lewis, J. y M. Springer** (2009), *Effective Technical Assistance Principles: Lessons from Three Performance Pay Programs*, Center for American Progress, Washington, DC.
- Loeb, S., L. Darling-Hammond y J. Luczak** (en prensa), "How Teaching Conditions Predict Teacher Turnover in California Schools", en *Peabody Journal of Education*.
- Malen, B.** (1999), "On Rewards, Punishments, and Possibilities: Teacher Compensation as an Instrument for Education Reform", en *Journal of Personnel Evaluation in Education*, vol. 12, núm. 4, pp. 387-394.
- McEwan, P.J.** (2008), "Quantitative Research Methods in Education Finance and Policy", en H.F. Ladd y E.B. Fiske (eds.), *Handbook of Research in Education Finance and Policy*, Routledge, Nueva York, pp. 87-104.
- McKinsey & Company** (2007), *How the World's Best-Performing School Systems Come Out on Top*, Identifying Teacher Quality Project, Washington, DC.
- Milanowski, A.T.** (2007), "Performance Pay System Preferences of Students Preparing to be Teachers", en *Education Finance and Policy*, vol. 2(2), pp. 111-132.
- Milgrom, P. y J. Roberts** (1992), *Economics, Organization & Management*, Prentice Hall, Englewood Cliffs, California.
- Muralidharan, K. y V. Sundararaman** (2008), *Teacher Incentives in Developing Countries: Experimental Evidence from India*, documento de trabajo 2008-13, National Center on Performance Incentives.

- Muralidharan, K. y V. Sundararaman** (2009), *Teacher Performance Pay: Experimental Evidence from India*, documento de trabajo 15323, National Bureau of Economic Research, consultado el 15 de diciembre de 2009 en www.nber.org/papers/w15323.pdf.
- Odden, A. y C. Kelley** (1997), *Paying Teachers for What They Know and Do: New and Smarter Compensation Strategies to Improve Schools*, Corwin Press, Thousand Oaks, California.
- OCDE** (2005), *Teachers Matter: Attracting, Developing and Retaining Effective Teachers*, OCDE, París.
- OCDE** (2007), *Education at a Glance 2007: OECD Indicators*, OCDE, París.
- OCDE** (2009a), *Evaluating and Rewarding the Quality of Teachers: International Practices*, OCDE, París.
- OCDE** (2009b), *Creating Effective Teaching and Learning Environments: First Results from TALIS*, OCDE, París.
- OCDE** (2010a), *Education at a Glance 2010: OECD Indicators*, OCDE, París, consultado en www.OECD.org/document/52/0,3343,en_2649_39263238_45897844_1_1_1_1,00.html.
- OCDE** (2010b), *La medición del aprendizaje de los alumnos: Mejores prácticas para evaluar el valor agregado de las escuelas*, OCDE, París.
- OCDE Steering Group on Teacher Evaluation and Incentive Policies** (2009), *Preliminary Advice on the Design of State-Level Pilots of Teacher Performance and Incentive Mechanisms*, OCDE, París.
- Patton, M.Q.** (2002), *Qualitative Research and Evaluation Methods*, 3a ed., SAGE, Thousand Oaks, California.
- Pfeffer, J. y N. Langston** (1993), "The Effect of Wage Dispersion on Satisfaction, Productivity, and Working Collaboratively: Evidence from College and University Faculty", en *Administrative Science Quarterly*, vol. 38, pp. 382-407.
- Podgursky, M. y M. Springer** (2007), "Teacher Performance Pay: A Review", en *Journal of Policy Analysis and Management*, vol. 26, núm. 4, pp. 909-949.
- Raue, K., K. MacAllum y L. Ristow** (2008), *Building Systems to Recognize Teachers of Excellence: Lessons from the Ohio Teacher Incentive Fund*, Westat, Rockville, Maryland.
- Raue, K., K. MacAllum, A. Winkler y L. Ristow** (2008), *Different Designs, Common Paths: A First Look at the Ohio Teacher Incentive Fund*, documento especial, Westat's Education Studies Group.
- Rogers, H.** (2009), "Piloting and Evaluating Teacher Incentive Programs", presentación en el Taller OCDE-México sobre los Estándares de Calidad Educativa y de Evaluación, Ciudad de México, julio.
- Rogers, F.H. y E. Vegas** (2009), *No More Cutting Class? Reducing Teacher Absence and Providing Incentives for Performance*, World Bank Policy Research, documento de trabajo núm. 4847, The World Bank, Washington, DC.
- Rose, S.** (2010), *Pay for Performance Proposals in Race to the Top Round II Applications*, memorándum, Education Commission of the States.
- Rossi, P.H., M.W. Lipsey y H.E. Freeman** (2004), *Evaluation: A Systematic Approach*, 7a ed., SAGE, Thousand Oaks, California.
- Roza, M. y R. Miller** (2009), *Separation of Degrees: State-by-State Analysis of Teacher Compensation for Master's Degrees*, Center on Reinventing Public Education Rapid Response Brief, Seattle, Washington.
- Sanders, W.L. y J.C. Rivers** (1996), *Cumulative and Residual Effects of Teachers on Future Student Academic Achievement*, Value-Added Research and Assessment Center, University of Tennessee, Knoxville, Tennessee.
- Sclafani, S. y M. Tucker** (2006), "Teacher and Principal Compensation: An International Review", Center for American Progress, Washington, DC, www.americanprogress.org/issues/2006/10/teacher_compensation_html/pdf/education_report.pdf, consultado en mayo de 2009.
- Shavelson, R.J. y L. Towne** (eds.) (2002), *Scientific Research in Education*, National Academy Press, Washington, DC.
- Skinner, B.** (1981), "Selection by Consequences", en *Science*, vol. 213, pp. 501-514.
- Slotnick, W.J.** (2009), *It's More than Money: Making Performance-Based Compensation Work*, Center for American Progress, Washington, DC.

- Snowden, J.** (2007), *The Future of Teacher Compensation: Déjà Vu or Something New?*, Center for American Progress, Washington, DC.
- Springer, M.** (2010), *Performance Incentives: Their Growing Impact on K-12 Public Education*, Brookings Institution Press, Washington, DC.
- Springer, M.** y **M. Winters** (2009), *New York City's School-Wide Bonus Pay Program: Early Evidence from a Randomized Trial*, documento de trabajo 2009-02, National Center on Performance Incentives.
- Springer, M., J. Lewis, M. Podgursky, M. Ehlert, L. Taylor, O. López y A. Peng** (2009), *Governor's Educator Excellence Grant (GEEG) Program Year Two Evaluation Report*, National Center on Performance Incentives, Nashville, Tennessee.
- Springer, M., J. Lewis, M. Podgursky, M. Ehlert, L. Taylor, O. López y A. Peng** (2009), *Governor's Educator Excellence Grant (GEEG) Program Year Three Evaluation Report*, National Center on Performance Incentives, Nashville, Tennessee.
- Springer, M., M. Podgursky, J. Lewis, M. Ehlert, T. Gronberg, L. Hamilton, D. Jansen, O. López, A. Peng, B. Stecher y L. Taylor** (2009), *Texas Educator Excellence Grant (TEEG) Program: Year Two Evaluation Report*, National Center on Performance Incentives, Nashville, Tennessee.
- Taylor, L., M. Springer y M. Ehlert** (2009), "Characteristics and Determinants of Teacher-Designed Pay for Performance Plans: Evidence From Texas' Governor's Educator Excellence Grant Program", en M. G. Springer (ed.) (2010), *Performance Incentives: Their Growing Impact on American K-12 Education*, Brookings Institution Press, Washington, DC.
- Tukey, J.W.** (1977), *Exploratory Data Analysis*, Addison-Wesley, Reading, Massachusetts.
- United States Department of Education** (2010), *Application for the Teacher Incentive Fund*, CDFA núm. 84.385, United States Department of Education, Washington, DC.
- Vegas, E. e I. Umansky** (2005), *Improving Teaching and Learning through Effective Incentives: What Can We Learn From Education Reforms in Latin America?*, The World Bank, Washington, DC.
- Watson, J.G., S.B. Kraemer y C.A. Thorn** (2009), *Data Quality Essentials: Guide to Implementation, Resources for Applied Practice*, Center for Educator Compensation Reform.
- Zenger, T. y C. Marshall** (1995), "Does Size Matter in Group Rewards? Factors Affecting the Incentive Intensity and Performance of Group-Based Pay Plans", documento presentado en la reunión anual de la Academy of Management, Vancouver, Columbia Británica.

Conclusión

Este informe presenta las principales conclusiones y recomendaciones del Consejo Asesor sobre Políticas de Evaluación e Incentivos Docentes que vio la luz gracias al Acuerdo de Cooperación entre el gobierno de México, representado por la SEP (Secretaría de Educación Pública), y la OCDE (Organización para la Cooperación y el Desarrollo Económicos). Este reporte forma parte de un conjunto de trabajo más amplio que incluye: tres publicaciones relacionadas,¹ documentos especializados, talleres y reuniones técnicas con funcionarios de la Secretaría de Educación Pública (SEP) e importantes actores implicados. Gran parte del trabajo se efectuó de manera conjunta con funcionarios de la SEP, cuyo secretario participó varias veces tanto en reuniones como en talleres.

En lugar de adoptar un enfoque temático o específico al país, la estructura y el contenido del informe reflejan la naturaleza interconectada y la perspectiva sistémica inherentes a las acciones de reforma a gran escala. El Consejo Asesor se centró en la elaboración de políticas públicas eficaces y oportunas para incentivar a los maestros en servicio, esto los condujo naturalmente a enfocarse en los marcos de evaluación docente y en la elaboración de estándares de enseñanza. Apoyándose en análisis de políticas públicas y procesos de reforma de distintos países, el Consejo Asesor elaboró un marco unificado para analizar y considerar las diversas prioridades e implicaciones de las políticas públicas. El capítulo 2 de este informe se centra en la manera en que las instituciones nacionales, en este caso la SEP, pueden llevar más allá las acciones de reforma si lo hacen de manera sistemática y metódica, dando seguimiento a las recomendaciones claves de la OCDE presentes en el *Marco de Políticas Públicas para la Implementación de Reformas Educativas*.

Respecto al Marco de Políticas Públicas, se hace una distinción entre las fases de diseño, planeación, implementación y evaluación de las políticas. El informe destaca la importancia de los conocimientos locales para complementar la investigación de la OCDE, junto a la experiencia y las prácticas internacionales. Esto implica reunir la información necesaria a nivel local, nacional e internacional para tomar decisiones informadas sobre el diseño de las políticas públicas, tomando en cuenta las limitantes y las oportunidades locales. El propósito del Marco de Políticas Públicas es guiar el proceso analítico que garantice una revisión sistemática y exhaustiva de las diferentes opciones de políticas públicas disponibles por parte de los actores pertinentes. Esto será crucial para adaptar las mejores prácticas y adecuarlas a los contextos locales. El mayor desafío es reunir distintos tipos de conocimiento – científico, técnico, profesional, administrativo, político y pedagógico – en los momentos adecuados. Es claro que los procesos de reforma en México requerirán un esfuerzo consecuente y sostenido.

Si bien la rendición de cuentas está en el centro del debate sobre las reformas a gran escala de las políticas públicas en muchos países, muy pocas veces se enuncia en México la noción de que todos los actores implicados deben ser directamente responsables de los resultados educativos. La discusión en este rubro gira en torno a las relaciones que existen entre los diversos actores en los diferentes niveles con el fin de determinar quién debe rendir cuentas a quién, y por qué. En México, la rendición de cuentas, y los estándares que la acompañan, sentaría las bases para apoyar de manera sostenida las reformas con el fin de mejorar el logro de los alumnos, el desempeño escolar y las prácticas de enseñanza.

El capítulo sobre rendición de cuentas pasa revista a algunos mecanismos conocidos en México y otros países que han planteado reformas a nivel del sistema para mejorar el logro de los alumnos. El capítulo sugiere que la rendición de cuentas basada en el establecimiento de expectativas claras, formuladas en estándares claros y ambiciosos, es un instrumento eficaz para la transparencia y el compromiso público de todos los actores. El informe sugiere también que lo ideal es emplear múltiples medidas de desempeño con referencias cruzadas, válidas y confiables, incluyendo los datos de las evaluaciones de los alumnos. Así, el informe brinda en el capítulo 4 una revisión profunda de las fortalezas, las debilidades y las oportunidades necesarias para un mayor perfeccionamiento de la evaluación ENLACE.²

A lo largo de esta obra se observa un hilo conductor: el aprendizaje y el crecimiento de los alumnos conforme pasa el tiempo son los criterios claves de los que deben rendir cuentas todos los actores del sistema educativo. Por la importancia que tienen los maestros y las escuelas, una unidad básica de rendición de cuentas (aunque no la única) debe ser la escuela. Todo alumno debe tener acceso a oportunidades significativas de aprendizaje de alta calidad. Es claro que el motor de las políticas para mejorar la educación es el aprendizaje del alumno, lo cual requiere el compromiso de los maestros en lo individual y una responsabilidad social compartida por parte de los actores implicados.

Medir el aprendizaje de los alumnos no es una empresa sencilla, y este informe reconoce la necesidad de elaborar más instrumentos para compilar más evidencias y validar los resultados de distintos enfoques experimentados. El capítulo 4 sugiere que, para complementar las evaluaciones estandarizadas o basadas en la escuela, también deben considerarse otros tipos de evaluaciones. Una de estas opciones se revisa en el capítulo 5 dedicado a la evaluación del valor agregado de las escuelas, tema en el que los países muestran cada vez más un mayor interés en años recientes. El diseño adecuado de los métodos de valor agregado puede producir mediciones del crecimiento del alumno más precisas y, a final de cuentas, más justas que las que se obtienen mediante comparaciones de promedios de puntuaciones brutas. Sin embargo, los desafíos son considerables, por lo que el capítulo bosqueja los problemas clave asociados a los modelos de valor agregado.

La opinión general reflejada en este informe es que el sistema educativo en su conjunto (autoridades, escuelas, maestros, padres de familia y alumnos) debe rendir cuentas del aprendizaje. Debido a la importancia que tienen los maestros, el capítulo 6 presenta directrices para la elaboración de un sistema de evaluación de maestros en servicio completo, justo y transparente. Esta propuesta reposa en particular en la elaboración de estándares claros de buenas prácticas de enseñanza.³ Si bien México ya ha tenido experiencias sobre la evaluación docente, cada vez hay más conciencia de la necesidad de mejorar estas experiencias hasta obtener un sistema sólido de evaluación docente. El desafío reside en conseguir que se dé la transición del sistema previo a uno nuevo y eficaz.

Bastante vinculado a un proceso de evaluación de los maestros de este tipo se encuentra el asunto de los incentivos docentes. El capítulo sobre incentivos para maestros en servicio revisa varios ejemplos internacionales e identifica las mejores prácticas en algunas áreas, como son los beneficios de los programas piloto y la manera de efectuar un monitoreo y evaluación adecuados de los programas y políticas de incentivos (véase el Apéndice 7A).

Los análisis, conclusiones y recomendaciones de este informe deben promover los procesos que se necesitan para alcanzar una implementación de las reformas educativas específica al país, apelando a análisis adicionales y a la creación de consensos a niveles federal, estatales y locales. La combinación en México de las prácticas internacionales y las limitantes de las políticas locales y específicas hace que la labor implícita en este informe haya sido tan desafiante como gratificante. El Acuerdo de Cooperación se benefició de los conocimientos de la OCDE, y empiezan a definirse ciertas estrategias para llegar a un compromiso eficaz. Este tipo de saber, junto con una perspectiva internacional comparativa, es lo que la OCDE seguirá poniendo al servicio de México y otros países miembros y asociados.

NOTAS

1. Las dos publicaciones hermanas son *Mejorar las escuelas: Estrategias para la acción en México* (OECD, 2010) y *Evaluación y reconocimiento de la calidad de los docentes: Prácticas internacionales* (OECD, 2009). Dentro del marco del Acuerdo de Cooperación se actualizó y tradujo al castellano un material importante sobre modelos de valor agregado: *La medición del aprendizaje de los alumnos: Mejores prácticas para evaluar el valor agregado de las escuelas* (OECD, 2011). Los documentos de trabajo encargados por la OCDE a expertos en el tema están disponibles en www.oecd.org/edu/calidadeducativa.
2. Evaluación Nacional del Logro Académico en Centros Escolares.
3. La publicación hermana *Mejorar las escuelas: Estrategias para la acción en México* (2010) presenta un análisis sobre la selección de maestros, así como la formación inicial y continua en México.

ORGANIZACIÓN PARA LA COOPERACIÓN Y EL DESARROLLO ECONÓMICOS

La OCDE constituye un foro único en su género, donde los gobiernos trabajan conjuntamente para afrontar los retos económicos, sociales y medioambientales que plantea la globalización. La OCDE está a la vanguardia de los esfuerzos emprendidos para ayudar a los gobiernos a entender y responder a los cambios y preocupaciones del mundo actual, como el gobierno corporativo, la economía de la información y los retos que genera el envejecimiento de la población. La Organización ofrece a los gobiernos un marco en el que pueden comparar sus experiencias políticas, buscar respuestas a problemas comunes, identificar buenas prácticas y trabajar en la coordinación de políticas nacionales e internacionales.

Los países miembros de la OCDE son Alemania, Australia, Austria, Bélgica, Canadá, Chile, Corea, Dinamarca, España, Estados Unidos de América, Estonia, Finlandia, Francia, Grecia, Hungría, Irlanda, Islandia, Israel, Italia, Japón, Luxemburgo, México, Noruega, Nueva Zelanda, Países Bajos, Polonia, Portugal, Reino Unido, República Checa, República de Eslovenia, República Eslovaca, Suecia, Suiza y Turquía. La Comisión de la Comunidad Europea participa en el trabajo de la OCDE.

Las publicaciones de la OCDE aseguran una amplia difusión de los trabajos de la Organización. Éstos incluyen los resultados de la compilación de estadísticas, los trabajos de investigación sobre temas económicos, sociales y medioambientales, así como las convenciones, directrices y los modelos desarrollados por los países miembros.

Establecimiento de un marco para la evaluación e incentivos docentes

CONSIDERACIONES PARA MÉXICO

En esta época marcada por las economías del conocimiento y cambios demográficos significativos, todos los sistemas educativos se encuentran ante la necesidad de mejorar los resultados del aprendizaje de sus alumnos y lograrlo aun con recursos limitados.

Con el fin de asistir a México y a otros países en este desafío, esta publicación presenta recomendaciones para el diseño, planeación, ejecución y monitoreo de políticas educativas en torno a la medición del aprendizaje, los estándares educativos y la evaluación docente, con base en las mejores prácticas internacionales. Considerando que la calidad de los resultados de los sistemas educativos no puede ser mayor que la calidad de sus docentes, esta publicación subraya la importancia de la evaluación y el reconocimiento de los docentes.

Sin embargo, la implementación efectiva de reformas educativas puede ser un proceso complejo y con grandes retos. La simple identificación y enumeración de las políticas que deberán ser implementadas no resulta suficiente. Los gobiernos también deberán determinar el “cómo” para el diseño y puesta en marcha de políticas educativas. Para ello, esta publicación presenta recomendaciones específicas y un método para que los tomadores de decisiones y actores relevantes analicen y se apropien de las mejores prácticas internacionales, pero adaptadas a los contextos locales.

Este reporte presenta las recomendaciones desarrolladas por el *Consejo Asesor de la OCDE sobre Políticas de Evaluación e Incentivos Docentes en México*, integrado por expertos reconocidos y analistas de la OCDE.

Publicaciones relacionadas

La medición del aprendizaje de los alumnos: Mejores prácticas para evaluar el valor agregado de las escuelas (OCDE, 2011)

Mejorar las escuelas: Estrategias para la acción en México (OCDE, 2010)

Evaluación y reconocimiento de la calidad de los docentes: Prácticas internacionales (OCDE, 2009)

Esta publicación es un producto del acuerdo de cooperación establecido entre el gobierno de México y la OCDE, el cual tiene como objetivo mejorar la calidad de la educación en México.

Por favor, cite esta publicación de la siguiente manera:

OCDE (2011), *Establecimiento de un marco para la evaluación e incentivos docentes: Consideraciones para México*, OECD Publishing.

<http://dx.doi.org/10.1787/9789264059936-es>

Este trabajo está publicado en *OECD iLibrary*, plataforma que reúne todos los libros, publicaciones periódicas y bases de datos de la OCDE. Visite www.oecd-ilibrary.org y no dude en contactarnos para más información.

2011

OECD publishing
www.oecd.org/publishing

ISBN 978-92-64-05999-3
87 2010 03 4P



9 789264 059993