

La evaluación de docentes de educación básica

Una revisión de la experiencia internacional

Felipe Martínez Rizo



Otros textos
de evaluación

INEE

Instituto Nacional para la
Evaluación de la Educación

México

La evaluación de docentes de educación básica

Una revisión de la experiencia internacional

LA EVALUACIÓN DE DOCENTES DE EDUCACIÓN BÁSICA.
UNA REVISIÓN DE LA EXPERIENCIA INTERNACIONAL

Primera edición, 2016
ISBN: 978-607-7675-83-9

Autor
Felipe Martínez Rizo

D. R. © Instituto Nacional para la Evaluación de la Educación
Av. Barranca del Muerto 341, Col. San José Insurgentes,
Delegación Benito Juárez, C.P. 03900, Ciudad de México

Coordinación editorial
María Norma Orduña Chávez

Corrección de estilo
Carlos Garduño González
Hugo Soto de la Vega

Formación
Heidi Puon Sánchez

Impreso y hecho en México.
Distribución gratuita. Prohibida su venta.

Consulte el catálogo de publicaciones en línea:

www.inee.edu.mx

La coordinación de esta obra estuvo a cargo de la Dirección General
de Difusión y Fomento de la Cultura de la Evaluación

El contenido, la presentación, así como la disposición en conjunto y de cada
página de esta obra son propiedad del Instituto Nacional para la Evaluación
de la Educación. Se autoriza su reproducción parcial o total por cualquier
sistema mecánico o electrónico para fines no comerciales y citando la fuente
de la siguiente manera:

Martínez Rizo, F. (2016). *La evaluación de docentes de educación básica.
Una revisión de la experiencia internacional*. México: INEE.

La evaluación de docentes de educación básica

Una revisión de la experiencia internacional

Felipe Martínez Rizo



Otros textos
de evaluación

INEE
Instituto Nacional para la
Evaluación de la Educación
México

DIRECTORIO

JUNTA DE GOBIERNO

Sylvia Irene Schmelkes del Valle
CONSEJERA PRESIDENTA

Eduardo Backhoff Escudero
CONSEJERO

Gilberto Ramón Guevara Niebla
CONSEJERO

Margarita María Zorrilla Fierro
CONSEJERA

Teresa Bracho González
CONSEJERA

TITULARES DE UNIDAD

Francisco Miranda López
UNIDAD DE NORMATIVIDAD Y POLÍTICA EDUCATIVA

Jorge Antonio Hernández Uralde
UNIDAD DE EVALUACIÓN DEL SISTEMA EDUCATIVO NACIONAL

Agustín Caso Raphael
UNIDAD DE INFORMACIÓN Y FOMENTO DE LA CULTURA
DE LA EVALUACIÓN

Susana Justo Garza (encargada)
UNIDAD DE PLANEACIÓN, COORDINACIÓN
Y COMUNICACIÓN SOCIAL

Miguel Ángel de Jesús López Reyes
UNIDAD DE ADMINISTRACIÓN

Luis Felipe Michel Díaz
CONTRALOR INTERNO

José Roberto Cubas Carlín
COORDINACIÓN DE DIRECCIONES DEL INEE
EN LAS ENTIDADES FEDERATIVAS

Dirección General de Difusión
y Fomento de la Cultura de la Evaluación
Annette Santos del Real

Dirección de Difusión y Publicaciones
Alejandra Delgado Santoveña

Índice

- 9 Presentación
- 11 Introducción

13 **Capítulo 1** La evaluación de docentes del siglo XVIII al XXI

- 13 Las escuelas premodernas y sus maestros
- 15 Irrupción de la modernidad
- 16 De fines del XIX a mediados del XX
- 20 De mediados del siglo XX al XXI
- 22 Evaluación por directores e inspectores
- 24 Pruebas para maestros y sistemas de observación simples
- 26 Esquemas de pago por méritos
- 28 Sistemas de indicadores múltiples
- 32 Modelos de valor agregado

35 **Capítulo 2** En busca de mejores sistemas de evaluación

- 35 Evolución de los sistemas de inspectores
- 36 Investigaciones sobre la efectividad de los docentes
- 39 Acercamientos basados en interrogación
- 44 Acercamientos basados en observación
- 49 Acercamientos basados en análisis de productos

55 **Capítulo 3** La evaluación de los maestros y la calidad educativa

- 56 Piezas de un sistema integral de evaluación de maestros
- 57 Acreditación de programas de formación inicial
- 57 Licencias para el ingreso a la profesión
- 58 Certificación de niveles avanzados de desempeño
- 59 Retos para el desarrollo de buenos sistemas de evaluación
- 59 Desafíos conceptuales: ¿qué es un buen maestro?
- 63 Desafíos técnicos: ¿cómo medirlos?
- 67 Desafíos prácticos: ¿cómo implementar las medidas?

71 **Conclusiones**

75 **Apéndice** La evaluación de los docentes en Iberoamérica

87 **Referencias bibliográficas**

Presentación

Esta revisión de literatura sobre los enfoques que se han utilizado a lo largo del tiempo para evaluar a los maestros, paralelamente a la expansión de la educación básica en los sistemas educativos de muchos países, es una prolongación de mis trabajos previos sobre los acercamientos a la observación de la práctica docente, así como a la evaluación formativa en aula y a la evaluación educativa en general.

Una versión más breve de este trabajo constituye un capítulo del Informe 2013 de seguimiento de las Metas 2021 de la Organización de Estados Iberoamericanos para la Educación la Ciencia y la Cultura. Asimismo, algunos elementos de este texto fueron retomados de un trabajo previo (Martínez Rizo, 2012a y 2012b).

El apoyo que he recibido de mi esposa durante más de 40 años de matrimonio ha sido fundamental para mi trabajo. Además de la dicha de compartir gozos y penas, hijos y nietos, sin este apoyo no habría podido dedicarme con la intensidad con que lo he hecho a la docencia y la investigación, tanto en la Universidad Autónoma de Aguascalientes, desde 1974, como en el Instituto Nacional para la Evaluación de la Educación, entre 2002 y 2008. Sin embargo, el nombre de María Elena Fernández aparece en pocas de las dedicatorias de los libros que he publicado. Al dedicarle ahora este trabajo, quiero también reconocer expresamente el valor de su presencia a mi lado durante todos estos años.

Felipe Martínez Rizo

Aguascalientes, junio de 2013

Introducción

En muchos sectores de la población mexicana, como ocurre también en otros países, hay actualmente una gran preocupación por la calidad de la educación y, en particular, por la enseñanza que ofrecen las escuelas de nivel básico.

Esa inquietud es alimentada en parte por interpretaciones mal sustentadas de los resultados de evaluaciones internacionales como las que impulsa la Organización para la Cooperación y el Desarrollo Económicos (OCDE), conocidas por las siglas PISA (*Programme for International Student Assessment*). En muchas ocasiones los medios de comunicación difunden la información derivada de esas pruebas en una forma que no toma en cuenta sus alcances y límites, lo que causa que el público no especializado —que comprende a la mayoría de la población, incluyendo a maestros, funcionarios de ministerios de educación y legisladores— se forme una idea muy inexacta de la situación real de las escuelas.

Esa preocupación por la calidad educativa, por otra parte, en muchos casos —y sin duda en el de México— tiene fundamento real, sobre todo si se tienen presentes las exigencias que el contexto económico, político y social plantea al sistema educativo en los albores del siglo XXI.

A diferencia de lo que pasaba hace no mucho tiempo, en el nuevo contexto ya no es suficiente que sólo una minoría de la población tenga un buen nivel educativo, en tanto que para la mayoría era suficiente una escolaridad elemental; ahora es necesario que todos los futuros ciudadanos desarrollen competencias cognitivas y no cognitivas de complejidad considerable, lo que están lejos de conseguir los sistemas educativos de muchos países, entre los que se encuentra el de México.

Ahora bien, pocas ideas reúnen un consenso tan amplio entre los interesados en educación como la que se refiere a la importancia del maestro para alcanzar un buen nivel de calidad, pero identificar a los maestros eficaces, distinguiéndolos de manera confiable de los que no lo son, es una tarea aún no resuelta. Muestra evidente de lo anterior son las formas que se han utilizado a lo largo del tiempo para aceptar que una persona desempeñe funciones docentes, así como para reconocer a los que lo hagan de manera destacada.

Por lo anterior se entiende que una de las varias formas en que se manifiesta la preocupación general por la calidad de la educación consista, en particular, en un reclamo por que se establezcan sistemas adecuados de evaluación de maestros; con ello se espera asegurar que toda persona con un grupo escolar a su cargo tenga las competencias necesarias para desempeñar su tarea en forma adecuada, impidiendo que lleguen al aula personas no calificadas y, en su caso, separando del servicio docente a quienes no cumplan con los estándares de calidad.

La insatisfacción con los sistemas usuales de evaluación del trabajo docente es muy amplia. Dichos sistemas suelen definir escalas de remuneración con base en la escolaridad y la antigüedad en servicio de los maestros, así como en evaluaciones hechas por directores y supervisores. Hay coincidencia en el sentido de que esos elementos no son los mejores para distinguir a los buenos docentes de los

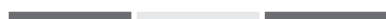
que no lo son, por lo que se buscan mejores acercamientos al estudio y la evaluación de la "efectividad" docente, tanto para sustentar formas justas de reconocer a los buenos docentes, como para identificar con precisión las necesidades de desarrollo profesional de todos y, llegado el caso, para tener bases sólidas que justifiquen retirar de la profesión a los que no deban estar en ella.

En el sistema educativo mexicano están presentes, con matices propios, tanto la preocupación general por la calidad, como el interés particular por la evaluación de los docentes; en este sentido, las formas de evaluar a los maestros y definir su salario son similares a las que se han usado en otros lugares.

Con el fin de contribuir a la reflexión respecto a cómo mejorar la calidad de la educación en nuestro país, este documento presenta un amplio panorama de la experiencia internacional sobre evaluación de maestros. ●

Capítulo 1

La evaluación de docentes del siglo XVIII al XXI



La situación de los maestros de los niveles escolares básicos se ha modificado a lo largo del tiempo, en relación con otros cambios de las escuelas y de la sociedad en general. Las formas en que los futuros maestros se preparaban para las tareas que desempeñarían, los procedimientos mediante los cuales eran admitidos en la profesión, y los que se utilizaban para decidir su permanencia en ella y la remuneración que recibirían son algunos de los aspectos que han cambiado a lo largo de la historia, a veces profundamente, a veces de manera superficial.

En este capítulo se describen las formas de seleccionar y evaluar a los maestros desde el siglo XVIII y hasta principios del XXI. Se deja para el capítulo siguiente la discusión de las pistas por las que transita la búsqueda actual.

LAS ESCUELAS PREMODERNAS Y SUS MAESTROS

En una perspectiva histórica, algunos rasgos clave de los sistemas educativos contemporáneos son relativamente recientes: los rasgos que permiten aplicar a los sistemas actuales el calificativo de “modernos”, en contraste con los que distinguían a la educación de épocas anteriores, son: universal frente a elitista; obligatoria frente a opcional; gratuita frente a pagada; laica, o al menos plural, frente a confesional.

No es difícil ver que esos rasgos de modernidad surgieron y tienen sentido en el contexto más amplio del desarrollo de las sociedades con la urbanización, la Revolución industrial y la economía capitalista; los sistemas políticos democráticos con división de poderes y separación de Estado e Iglesia, y las ideas de la Ilustración sobre los derechos civiles y la igualdad de los ciudadanos ante la ley.

Las sociedades del “antiguo régimen” se caracterizaban por los rasgos opuestos: producción agrícola y artesanal en un medio masivamente rural; economía feudal, absolutismo político; yuxtaposición Estado-Iglesia; aceptación de la desigualdad ante la ley, y existencia de privilegios de nacimiento.

En esas sociedades premodernas no era necesario que el grueso de la población supiera leer y escribir; bastaba con que pudiera hacerlo una minoría de clérigos

y unos profesionales más, como médicos y abogados. En las sociedades actuales, en cambio, es necesario que todo ciudadano, además de saber leer y escribir, tenga conocimientos básicos de Matemáticas y de Ciencias Naturales y Sociales.

El avance de la “sociedad del conocimiento” hace necesario que la población tenga “competencias” cada vez más complejas en esas y otras áreas, como una segunda o tercera lengua, computación, creatividad, trabajo en equipo, etcétera.

Las nuevas exigencias que trae consigo el desarrollo de las sociedades explican, pues, que se hayan creado sistemas escolares masivos, para atender a toda la población, y que haya sido el Estado —ahora desligado de la Iglesia— el que haya debido encargarse de ello, lo que implica que la educación sea, en general, gratuita y laica, como servicio público que es.

En la sociedad del “antiguo régimen”, en cambio, las escuelas de primeras letras atendían a una minoría de alumnos privilegiados cuyas familias debían pagar al maestro. En una época en que Iglesia y Estado estaban imbricados, las escuelas no conformaban un servicio público, y el oficio de maestro de primeras letras era controlado en los países europeos y sus colonias por la Iglesia y el gremio. En los siglos XVII y XVIII, para llegar a ser admitido en la corporación, un joven debía vivir en casa de “su” maestro, como aprendiz y sirviente, hasta saber lo necesario. En la Nueva España las ordenanzas de la Cofradía de San Casiano, que databan del siglo XVII, establecían como requisitos para ser aceptado:

- Demostrar “limpieza de sangre”, en otras palabras, ser de raza blanca pura: “el que hubiere de ser maestro no ha de ser negro ni mulato ni indio”.
- Demostrar ser “cristiano viejo”, lo que excluía a los judíos conversos, ya que los indígenas estaban descartados por el requisito anterior: “y siendo español ha de dar información de cristiano viejo de vida y costumbres”.

Sólo después veían los requisitos de conocimientos y habilidades:

- Saber leer y escribir y manejar la aritmética elemental: “ha de saber leer romance en libros y cartas misivas y procesos, y escribir las formas de letras siguientes: redondillo grande y más mediano y chico, bastardillo grande y más mediano y chico [...] y ha de saber [...] las cinco reglas de cuenta guarisma, que son sumar, restar, multiplicar, mediopartir y partir por entero” (Gonzalbo Aizpuru, 2001: 240-241).

Las mujeres, desde luego, simplemente no podían aspirar a ser maestras.

Las colonias de la Gran Bretaña en Norteamérica fueron establecidas, en general, por miembros de grupos religiosos disidentes en busca de aires más propicios que los que les ofrecía la Iglesia anglicana en la metrópoli. Su postura recelosa frente a todo Estado central se reflejó en el sistema federal que surgió tras la independencia de los Estados Unidos, en 1776.

Lo anterior, junto con la ausencia de un gremio fuerte en lo que se refiere a la organización de las escuelas elementales, explica que hubiera algunos rasgos similares a los que prevalecían en países con tradición central como

Francia, o asimismo España y sus colonias, pero también variantes peculiares. En un trabajo sobre la historia de la certificación de los maestros en los Estados Unidos, Sedlak señala que en las colonias de Norteamérica:

se solicitaba a actores locales diversos (religiosos veteranos, ciudadanos prominentes, comités de laicos) que seleccionaran y contrataran maestros, lo que hacían utilizando procedimientos informales para decidir sobre los prospectos: se basaban en recomendaciones de los ministros de culto; interrogaban a los candidatos sobre sus creencias y valores; valoraban su fortaleza física y su valor; o contrataban a sus parientes, a los que conocían muy bien (2008: 856, en Wilson, 2009: 16).

IRRUPCIÓN DE LA MODERNIDAD

Los cambios que dieron origen a los sistemas educativos modernos trajeron consigo la supresión de los gremios, incluyendo el de maestros, lo que implicó cambios sustanciales en cuanto a la formación de los futuros docentes y a las formas de evaluarlos para su ingreso y permanencia en la profesión.

Para la formación inicial surgieron “seminarios pedagógicos” (Prusia) y “escuelas normales” (Francia). Para la evaluación de docentes, los sistemas de inspectores.

La formación incluía el estudio de los pioneros de la pedagogía, como Comenio y luego Pestalozzi o Herbart, pero la atención se centraba en el desarrollo de la personalidad de los maestros como personajes cuya influencia sería fundamental para que los nacientes Estados tuvieran una ciudadanía que compartiera los valores de una modernidad —republicana o no— basada en el respeto de la ley y las instituciones.

Como apunta Glenn en la introducción de su importante estudio sobre la escuela pública, en el proyecto de ésta: “tenía mucho más importancia fundamental la ‘educación moral’ y la formación de una identidad nacional compartida que la transmisión de unos conocimientos académicos básicos” (2006: 19).

El nombre mismo de “seminarios pedagógicos” sugiere la idea de un sacerdocio laico. Con otra designación —que subraya la importancia de que la educación “uniformara” a los futuros ciudadanos—, las escuelas normales francesas tenían propósitos similares de los que se derivaba un modelo formativo que compartía también los rasgos de los seminarios católicos tridentinos, concebidos como instituciones que no buscaban ante todo preparar técnicos competentes, sino apóstoles consagrados a un ideal. Los nombres con los que se designaban los libros diseñados para servir como textos en las escuelas laicas surgidas de la Revolución francesa revelan esa similitud de propósitos: *catecismos*, *epístolas* y *evangelios republicanos* (cfr. Glenn, 2006: 41).

Por lo que se refiere al monitoreo del desempeño de los nuevos maestros, el mecanismo que los nuevos Estados desarrollaron, en ausencia del control por el gremio o la Iglesia, pero con claras raíces en las tradiciones correspondientes, fue el de los servicios de inspectores.

En 1802 Napoleón creó el primer cuerpo de inspectores para los liceos; en 1833 Guizot estableció la inspección para las escuelas primarias, y en 1852 Napoleón III

dio forma a la Inspección General de la Instrucción Pública. En España el Cuerpo de Inspectores de la Enseñanza en las Escuelas Primarias se creó mediante un decreto del 30 de marzo de 1849. En las antiguas colonias españolas de América se crearon cuerpos similares en distintas fechas comenzando, al parecer, en Costa Rica, desde la década de 1830 (SICI, 2013b; SICI, 2013c; Lillis, 1992: 4-5).

En el mundo anglosajón, por la misma época surgieron organismos similares con rasgos propios. En la Gran Bretaña, como la Iglesia anglicana tenía el control de la mayor parte de las escuelas, los primeros inspectores fueron designados por ella en 1837; se designaron luego inspectores para atender escuelas de otras denominaciones religiosas, hasta que en 1876 el Inspectorado de Su Majestad se reorganizó, con base en las divisiones políticas, para Inglaterra, Gales, Escocia e Irlanda (SICI, 2013d; Bolam, 1985: 4429-4432)

El modelo napoleónico de inspectorado influyó en los que se desarrollaron en las colonias francesas, pero también en el sistema español y en los que surgieron en las antiguas colonias de España en América, ya independientes, en las que el desarrollo de sistemas educativos modernos fue más lento que en las principales naciones europeas.

La inestabilidad derivada de la pugna entre liberales y conservadores que marcó la historia de las naciones de Iberoamérica durante todo el siglo XIX, muchas veces en formas más o menos abiertas de guerra civil, hizo que las instituciones de la modernidad educativa se implantaran lenta e irregularmente.

Las primeras escuelas normales de la región datan de la década de 1820, y, al parecer, en Costa Rica se creó un sistema de inspectores en la de 1830, antes incluso que en la Gran Bretaña (Lillis, 1992), pero en general hubo que esperar hasta las últimas décadas del XIX para que normales e inspectorados se consolidaran. Durante ese lapso, la autorización para trabajar como maestro en alguna escuela era otorgada por la autoridad municipal o regional, por la autoridad eclesiástica o por alguna combinación de una y otra, según el estatus del plantel de que se tratara.

Los modelos de inspectorado de Prusia, los Países Bajos y Gran Bretaña, por su parte, influyeron en los de las colonias que luego formarían la *Commonwealth*, en particular Canadá, Australia y Nueva Zelanda, así como en los Estados Unidos. El carácter descentralizado del sistema educativo de este último país, congruente con su carácter federal, hizo que los inspectorados dependieran de cada uno de los estados, con diferencias importantes entre unos y otros.

DE FINES DEL XIX A MEDIADOS DEL XX

El ideal ilustrado de que toda la población tuviera al menos la educación elemental no se hizo realidad en poco tiempo. En los países más pobres, incluidos algunos de América Latina, no se consigue todavía. Los primeros países que lograron la universalización efectiva de la enseñanza elemental —los Estados Unidos y algunas naciones de Europa— lo consiguieron sólo a fines del siglo XIX o principios del XX.

Fue la existencia efectiva de sistemas masivos de educación primaria lo que hizo que se modificaran la organización y la operación de las escuelas. El sistema

de aula única para todos los grados, por ejemplo, dejó el lugar a escuelas organizadas por grados; los inspectores también se diversificaron.

En paralelo, el avance en las concepciones ilustradas sobre la igualdad de derechos de todas las personas hizo que los criterios de raza o género, que tradicionalmente definían las remuneraciones, fueran sustituidos por sistemas basados en la escolaridad y la antigüedad, pero no acabó de golpe con las viejas ideas que estaban profundamente arraigadas en las estructuras sociales y en la mentalidad de las personas.

Ejemplo de ello es la persistencia de la esclavitud en los Estados Unidos, pese a que contradecía los ideales plasmados en los textos fundacionales de la nueva república. La abolición conseguida por Lincoln durante la Guerra Civil no fue suficiente para alcanzar la anhelada igualdad, y hubo que esperar hasta las décadas de 1950 y 1960 para que el movimiento por los derechos civiles lograra que se prohibiera la segregación racial en las escuelas.

A fines del siglo XIX la consolidación de sistemas de educación elemental fue acompañada por la estandarización de los salarios de los maestros, con base en el nivel en que enseñara cada uno, su antigüedad, su escolaridad y algún tipo de evaluación de su desempeño por parte de un supervisor, pero además según el género y la filiación étnica del docente. Se utilizaban, pues, “escalas salariales diferenciadas” (*tiered system*) (Gratz, 2009: 57). Según Gratz, en Boston, en 1876:

el salario de un maestro varón en una escuela elemental iba de 1700 a 3200 dólares, en tanto que el de una maestra iba de 600 a 1200 dólares; en una escuela secundaria un hombre ganaba de 1700 a 4000 dólares, y una mujer de 1000 a 2000 (2009: 58).

El avance de los derechos de las mujeres hizo que las diferencias salariales se redujeran. El derecho al voto quedó plasmado en la 19ª Enmienda constitucional en 1920, y las primeras “escalas únicas de salarios” (*single salary schedule*) en el sentido de una sola escala para hombres y mujeres fueron adoptadas en 1921 en Denver y Des Moines. Para 1925, “casi 80 por ciento de las maestras en las principales ciudades de la nación habían logrado tener un salario igual al de sus colegas varones por un trabajo igual” (Prostik, 1996: 275, en Gratz, 2009: 58).

La escala única de salarios buscaba, pues, “erradicar la práctica de pagar salarios diferentes a los maestros” según género y raza. En palabras de Prostik (1996):

Esta escala única utiliza las unidades y grados de escolaridad y los años de experiencia como base para pagar un salario diferente a cada maestro. La idea subyacente es que un maestro con más escolaridad y antigüedad tiene más competencia que los demás y, por consiguiente, debe recibir mejor remuneración (en Porter, Youngs y Odden, 2001: 288).

Aunque las normas hubieran cambiado, la discriminación basada en criterios de raza llevó más tiempo para desaparecer que la derivada del género, sobre todo en los lugares en que la mentalidad racista estaba más arraigada. En las primeras décadas del siglo XX, en los estados sureños, se exigía que los maestros

afroamericanos fueran evaluados anualmente para renovar sus contratos, con evaluaciones muchas veces arbitrarias y caprichosas.

El historiador Van Littlefield (1994) documenta la manera en que los miembros blancos de los comités escolares de evaluación iban a las aulas de maestras afro-americanas y decidían sobre la calidad de su trabajo para poder renovar sus contratos. Un número desproporcionado de esas decisiones se hacía debido a la aceptación, por esas mujeres, de acciones de acoso sexual (Ladson-Billings, 2009: 215).

Para mediados del siglo XX casi todas las escuelas de los Estados Unidos habían adoptado la escala única de salarios (Sharpes, 1987, en Odden y Kelly, 2002: 32). En las principales naciones europeas y en otros países en que el sistema político adoptó los rasgos de un Estado moderno, los viejos criterios que consideraban las características ideológicas, religiosas y raciales dejaron el lugar a sistemas de remuneración de maestros o escalafones basadas en escolaridad y antigüedad como los de cualquier funcionario.

Por lo que se refiere a los cuerpos de inspectores, en los países en que éstos se habían constituido en aparatos de Estado fuertes, con grados considerables de independencia respecto a los ministerios de educación, la masificación de los sistemas educativos hizo que un cuerpo nacional fuera insuficiente para atender un gran número de escuelas, por lo que se crearon instancias similares a nivel subnacional o para sectores especiales del sistema educativo.

La desconfianza frente a la intervención gubernamental en cualquier ámbito de la vida hizo que en los Estados Unidos la función de inspección no diera lugar a un aparato estatal independiente, sino que se encomendara a personal que, como los maestros, dependía de la autoridad de cada distrito escolar, la cual hasta fines del siglo XIX, por lo general estaba a cargo de consejos (*boards*) de ciudadanos.

La "ciudadanización" tiene ventajas, pero también desventajas que a fines del siglo XIX se manifestaban, en particular, en lo relativo a la selección y la promoción de los maestros. Un informe de 1890 dice:

Si el sistema educativo norteamericano quiere enfrentar con éxito sus retos actuales [...] necesita contar con un personal docente más sólido. Tal vez sólo un maestro de cada cinco o cada cuatro es un profesional [...] Me dirán ustedes que la ley regula eso [...] pero ¿qué es lo que hace la ley? Por lo general confiere la autoridad de certificar a los maestros, a los miembros del consejo de educación de la ciudad o el condado, o al comisionado del distrito, pero los miembros del consejo no son profesionales de la educación. ¿Cómo decidirán de manera inteligente quién está calificado para enseñar en una escuela?

Pero eso no es lo peor, porque si quisieran podrían contratar a un profesional para esas decisiones. Los miembros del consejo tienen la autoridad para contratar maestros, tienen tíos y primos, hijas y sobrinas que quieren trabajar. También tienen amigos personales y aliados políticos con sus respectivos parientes, amigos y conocidos. Son humanos. Quieren agradar. Solamente los más fuertes de ellos se atreven a enfrentar las incomprendiones y enemistades en las que se verán envueltos si se niegan a ayudar a sus amigos (Blodgett, 1893: 1163, en Rivkin, 2005: 257).

La insatisfacción con los sistemas prevalecientes se combinó con el surgimiento de nuevas formas de organizar y controlar el trabajo en las industrias, de las que se derivaron aumentos impresionantes en la productividad, por lo que comenzaron a influir en otros ámbitos de la vida social, como la educación.

Durante las primeras décadas del siglo XX en los Estados Unidos se desarrollaron las formas modernas de organización del trabajo industrial con el taylorismo y las cadenas de producción de Sears & Roebuck y Henry Ford. En estos esquemas productivos el papel del inspector era importante a fin de asegurar que todos los trabajadores siguieran fielmente las instrucciones para llevar a cabo su tarea de la manera más eficiente.

En un contexto en el que la mayoría de los docentes eran mujeres, en buena parte de origen afroamericano, que tenían una preparación deficiente, fueron bien recibidas las ideas de reformadores que, ante la dificultad de volver muy competentes a miles de docentes mal preparados, centraron sus esfuerzos en promover innovaciones curriculares, plasmarlas en libros de texto detallados, y hacer que los inspectores se aseguraran de que los docentes siguieran al pie de la letra las indicaciones del texto (Bolin y Panaritis, 1992).

Por otra parte, desde fechas tempranas fue claro que la evaluación por sí misma no mejora la calidad, por lo que a la función evaluadora y de control administrativo del trabajo escolar que denota el término *inspección* convenía añadirle una función paralela de apoyo técnico-pedagógico para promover mejoras, la cual busca denotarse con el término *supervisión*.

Sin embargo, las exigencias de la función de ayuda para la mejora pueden entrar en conflicto con las de control de calidad, ya que para brindar apoyo a un maestro éste deberá dejar que se manifiesten los puntos débiles de su trabajo, lo que puede tener consecuencias negativas en lo relativo a evaluación.

Por ello la función de apoyo se puede manejar como parte de la tarea de los inspectores o asignarse a otras instancias; manejar juntas o separadas las dos funciones tiene ventajas y desventajas, y los organismos que deben desarrollar las funciones gemelas de control y mejoramiento de la calidad, de inspección y supervisión pueden ser uno solo o dos; pueden ser independientes del Ministerio de Educación o depender de él, lo que parece más frecuente (Lillis, 1992: 1).

Por todo lo anterior, a lo largo del siglo XX los sistemas de inspectores se fueron desarrollando en formas similares pero no idénticas en los diversos sistemas educativos, lo que dio lugar progresivamente a estructuras complejas de instancias nacionales y subnacionales más o menos diversificadas y especializadas.

Según Watson, con la excepción de naciones muy pequeñas:

generalmente hay un cuadro de inspectores nacionales, federales o estatales, según los patrones de la administración de cada país, junto con un cuadro mucho mayor de inspectores regionales, departamentales o locales, muchos de los cuales son asesores y supervisores, tanto y más que inspectores propiamente dichos. En todos los casos pueden tener responsabilidad de áreas académicas específicas o de niveles particulares del sistema educativo (1995: 5247).

En los sistemas educativos de la mayor parte de los países europeos los cuerpos nacionales o regionales de inspectores siguen siendo una pieza clave del control de la calidad educativa. Cuerpos similares funcionan también en los sistemas educativos de países de Asia.

En las antiguas colonias de España en Latinoamérica y en las de Francia en África siguen funcionando sistemas de inspectores, aunque en ocasiones de manera muy ineficiente (Watson, 1995: 5247).

DE MEDIADOS DEL SIGLO XX AL XXI

Hoy está claro que los sistemas basados en escolaridad y antigüedad tampoco bastan para distinguir los mejores docentes de los menos buenos, pero conviene no olvidar que, en su momento, las escalas de salario basadas en tales criterios fueron un avance en dirección de la racionalidad weberiana, frente a los privilegios de los gremios y otros sectores en el antiguo régimen.

La insuficiencia de esos criterios para distinguir a los mejores maestros de los menos buenos llevó, a partir de la segunda mitad del siglo XX, a la búsqueda de mejores formas de evaluación de los docentes.

Un poderoso estímulo adicional se derivó, una vez más, de modificaciones de los sistemas educativos, relacionadas a su vez con cambios sociales más amplios: los que tienen que ver con el avance de las ciencias y las tecnologías, que ha dado lugar a las sociedades del conocimiento.

En estas sociedades, cada vez más globalizadas, la competencia económica hace que la competitividad del aparato productivo de un país sea fundamental, lo que trae consigo la necesidad de una fuerza de trabajo cada vez más preparada. Las evaluaciones internacionales, que buscan medir el nivel medio de competencia de los estudiantes de diversos países, han adquirido así gran visibilidad.

Todo esto tiene importantes implicaciones para los sistemas educativos: cuando su papel consistía en preparar una minoría muy calificada y una mayoría con niveles sólo elementales para la realización de trabajos rutinarios, las exigencias para alumnos y maestros eran muy distintas de las que trae consigo el propósito de preparar a todos los futuros ciudadanos para que desempeñen actividades que implican competencias cognitivas y no cognitivas más complejas. Como dice Linda Darling-Hammond:

A medida que los propósitos de la educación cambian del aprendizaje de datos y habilidades básicas al desarrollo de competencias de pensamiento y desempeño de alto nivel, cambian también las concepciones de lo que los maestros necesitan saber y poder hacer (2008: 761).

Las implicaciones para los sistemas de evaluación de maestros son claras: si se pretende que los alumnos aprendan cosas simples, la enseñanza también se podrá sistematizar en forma simple, identificando tareas que el docente deberá aplicar de manera rutinaria, y siguiendo al pie de la letra el libro de texto y otras ayudas curriculares. Si se trata de desarrollar competencias complejas, en

cambio, los docentes deberán identificar las necesidades particulares de alumnos diferentes y diseñar e implantar procesos de enseñanza personalizados.

En un caso se verá a los maestros como técnicos, en el otro como profesionales. En un caso se les podrá evaluar mediante pruebas con preguntas de respuesta estructurada o listas de cotejo para verificar que ocurran conductas previamente determinadas; en el otro serán necesarias formas de evaluación más complejas.

La insatisfacción con las escalas de salarios basadas en antigüedad y escolaridad se reflejó en los cuestionamientos presentados por el economista Mancur Olson en 1975, en una reunión organizada por el Educational Testing Service sobre los indicadores de calidad de los sistemas educativos. En esa ocasión Olson presentó unas provocadoras reflexiones sobre las razones por las que, en su opinión, no se suele contar con buenas mediciones de la calidad de los servicios públicos.

En el sector privado es usual que la tarea de identificar los mejores productos se deje al mercado, pero éste suele ser inadecuado para valorar la calidad de bienes públicos como los que tienen que ver con políticas de defensa o protección del ambiente. Para ilustrar su idea, Olson planteaba:

¿Se han preguntado ustedes por qué se espera que los oficiales del ejército tengan el calzado especialmente lustroso? Obviamente se puede pelear igual con zapatos sin lustrar, y en el fragor de la batalla se puede tolerar. Si mi punto de vista tiene aunque sea la mitad de la razón, la respuesta a esa pregunta tiene que ver con que el lustrado del calzado es una medida aproximada (proxy) de la dedicación al deber y la efectividad de los oficiales y, en épocas de paz, tal vez sea una aproximación tan buena como otras posiblemente disponibles (1975: 49).

Olson señalaba que las agencias que producen bienes públicos suelen tener sistemas de evaluación rudimentarios, con sistemas de servicio civil de carrera o méritos; credencialismo, y excesivo cuidado por apegarse a reglamentos y procedimientos, o para evitar problemas con los superiores. Y añadía:

Me parece que la razón de todo esto no es porque en esas instituciones haya personas que, por motivos perversos, usan formas inadecuadas de evaluar el desempeño; la razón, simple y sencillamente, es que en esas instituciones es especialmente difícil disponer de las medidas adecuadas. Por eso he dicho que se usa el criterio de las 4 "eses": Schooling (escolaridad), Seniority (antigüedad), Shoeshining (aseo del calzado) y Sycophancy (sicofancia, adulación) (50).

El análisis de Olson sugiere que hay razones técnicas que explican el uso de criterios similares a las 4 eses en sistemas tradicionales de evaluación como los que se utilizan para los maestros. Frente a los criterios prevalecientes en el sistema estadounidense en el siglo XIX, o los de la Cofradía de San Casiano, la escala única fue un avance, pero el uso de años de escolaridad o de experiencia como criterio tampoco es perfecto:

Las modalidades más usuales de la escala única no apoyan la necesidad de los maestros de ampliar o cambiar sus conocimientos y habilidades profesionales [...] la investigación muestra que, en promedio, después de unos cuantos años de trabajo docente una mayor experiencia no se asocia con más capacidad o más éxito del maestro en el aula (Prostik, 1996).

Evaluación por directores e inspectores

La idea de que directores e inspectores escolares pueden ser fuentes confiables de juicios de evaluación sobre el desempeño de los maestros a su cargo parece razonable, ya que el director —y, en menor medida, el supervisor— puede observar con frecuencia el trabajo de cada maestro, y se debería suponer que unos y otros cuentan con la formación y la experiencia necesarias para ser observadores altamente calificados. Por ello la evaluación de docentes se ha basado por lo general en la apreciación de esos funcionarios.

En la realidad en muchos casos la formación de directores y supervisores y la de los maestros tienen debilidades similares, y las tareas administrativas dejan poco tiempo para atender la función de apoyo que esos puestos deberían implicar. Por lo demás, la dimensión sumativa y la formativa de la evaluación —y los papeles de jefe y colega— entran en conflicto, ya que no es posible evitar la presencia de sesgos en sentido positivo o negativo en relación con los docentes.

A ello se añade que los instrumentos de observación de la práctica docente más frecuentemente utilizados consisten en listas de cotejo que sólo pueden captar la presencia o ausencia de conductas simples, como si fuera posible reducir a ellas la complejidad de una docencia profesional.

Peterson (2000: 28) señala la falta de sustento que tiene la fascinación por juicios evaluativos que se basan en instrumentos cuyo resultado final es una calificación numérica. El número da impresión de exactitud, pero ésta es falsa si se obtiene mediante una combinación arbitraria de puntuaciones parciales, cada una de las cuales se basa, a su vez, en escalas construidas en forma poco rigurosa.

En esa obra, Peterson cita una crítica a la “sobre-cuantificación” de los sistemas de evaluación de maestros que hace James Popham, para quien los números, desafortunadamente, no protegen contra la falta de sentido (Peterson, 2000: 28).

El resultado de todo ello es conocido: las evaluaciones de los docentes hechas por los directores o los inspectores escolares carecen de confiabilidad y, en muchos casos, presentan la característica de que prácticamente todos los evaluados obtienen resultados altos o, al menos, “satisfactorios”.

Según Peterson, “setenta años de investigación sobre la calificación de maestros por parte de directores muestran que no funciona bien”. Este investigador cita en apoyo de su dicho a Medley y Coker (1987: 242), quienes, tras revisar una docena de estudios al respecto llevados a cabo entre 1921 y 1959, muestran que todos llegaron a la misma conclusión: que “las correlaciones entre los puntajes con que los directores califican a los maestros y medidas directas de su efectividad se ubicaron alrededor de cero”. Después cita estudios más recientes, hasta 1990, que llegan a idéntico resultado (Peterson, 2000: 19-30).

La reducida variación de resultados de las evaluaciones de docentes a cargo de directores y supervisores, con proporciones casi nulas de puntajes insatisfactorios, se sigue presentando, aunque ya haya mejores protocolos de observación. En ese sentido ha llamado la atención un informe de Weisberg y sus colaboradores (2009), que reporta los hallazgos de un estudio sobre los sistemas de evaluación de docentes en 10 distritos escolares de los Estados Unidos, con más de 15 000 maestros.

Según ese informe, las normas que regulan la contratación de maestros suelen basarse en el falso supuesto “de que cualquier maestro puede llenar cualquier vacante, lo que lleva a las escuelas a contratar maestros que no quieren, y a los maestros a ocupar plazas para las que tal vez no sean adecuados” (2009: 9).

La expresión *widget effect*, que utilizan Weisberg y sus colaboradores, cuestiona la idea de que los profesores son “partes intercambiables” que pueden servir igual en cualquier escuela o contexto, como las piezas de una máquina, o un obrero de una cadena de producción, cuyo trabajo simple y rutinario permite que pueda ser remplazado por otro cualquiera en caso de necesidad.

Esta visión del trabajo docente recibe apoyo de sistemas de evaluación que no consiguen identificar docentes de diferente efectividad. Según Weisberg y sus colaboradores (2009: 11-12), en cinco distritos que utilizaban sólo dos categorías para calificar a los docentes, en prácticamente todos los casos (más de 99%) su desempeño fue considerado “satisfactorio”; en otros cinco distritos que utilizaban un número mayor de categorías para evaluar el desempeño docente, 70% de los maestros recibió la nota más alta, y otro 24%, la segunda más alta.

Esos sistemas de evaluación no detectan a los maestros no eficaces, y tampoco a los realmente sobresalientes, con obvias implicaciones tanto para propósitos sumativos (premios o sanciones), como para los formativos, ya que con base en esos resultados no es posible diseñar actividades de desarrollo profesional en función de necesidades de los docentes que presenten debilidades. En seis de los diez distritos escolares estudiados, en un lapso de cinco años no hubo un solo caso de maestro despedido por bajo rendimiento. En dos casos los despidos representaron 0.01%, en uno 0.04% y en otro 0.07% (Weisberg et al., 2009: 17).

En casi todos los casos la evaluación implica sólo una o, como máximo, dos visitas al aula, con una duración total de una hora o menos en 99% de los casos, y de media hora o menos en 50%. Además los directores reportan que no suelen recibir entrenamiento en el uso de protocolos de observación, sino que deben aprender a utilizarlos sobre la marcha (Weisberg et al., 2009: 21-22).

Son evidentes las limitaciones de esa forma de evaluar a un maestro con base en la apreciación de una sola persona, a partir de una visita de corta duración para observar el trabajo en aula, en muchos casos con anuncio previo, lo que permite al docente prepararse y mostrar el mejor desempeño de que es capaz, pero no el que lleva a cabo cotidianamente.

Un texto reciente retoma una expresión de uso común en los medios educativos, para caricaturizar lo que suele ocurrir en esas visitas de inspección: *dog & pony show*, en alusión a los espectáculos circenses en que un domador exhibe las habilidades de algunos animales bien entrenados. El autor propone sustituir

esas visitas únicas y anunciadas —a su juicio “imprecisas, deshonestas e ineficaces”— por “hasta 10 visitas cortas, no anunciadas, seguidas por retroalimentación oportuna y valiosa, cara a cara” (Marshall, 2012: 19-20).

Se puede pensar que la escasa variación de las calificaciones que resultan de las visitas al aula de directores e inspectores refleja el alto grado de uniformidad que pueden presentar las actuaciones breves y bien preparadas de muchos maestros, sobre todo si los aspectos de la docencia relevantes para la observación se refieren a conductas simples, que no es difícil atender, como comenzar a tiempo, anotar en el pizarrón el aprendizaje esperado de la sesión, hacer trabajar en grupo a los estudiantes, o utilizar una variedad de materiales instruccionales.

Sin llegar a extremos de corrupción como los que se han dado en algunos casos, esa forma de evaluar a los docentes, basada en el juicio inapelable de una persona, no es la mejor para identificar a los eficaces e ineficaces, sobre todo cuando el número de maestros que debe atender cada supervisor hace que visite cada aula sólo una vez en cada ciclo escolar, por poco tiempo y sin posibilidades de profundizar en el análisis de la práctica de cada docente.

Las dificultades asociadas con este enfoque se derivan de que es intensivo en mano de obra, basado en un periodo limitado de observaciones y vulnerable a los sesgos personales de los inspectores. Además, con frecuencia no es factible contar con un número suficiente de inspectores altamente calificados. Estas dificultades pueden explicar por qué el enfoque de la evaluación de los docentes basado en inspectores ha desaparecido casi totalmente en los Estados Unidos y Australia, donde antiguamente fue predominante, aunque sigue existiendo en muchos países europeos (Stufflebeam y Nevo, 1995: 2126).

Por esas limitaciones y por la mayor presencia de investigadores profesionales, entre otros factores, los sistemas de evaluación de maestros de los Estados Unidos han evolucionado más rápidamente que en otros países. En los puntos siguientes de este capítulo se sintetiza información sobre los avances al respecto en la segunda parte del siglo XX y principios del XXI.

Pruebas para maestros y sistemas de observación simples

Desde la década de 1970, los aspirantes a plazas docentes en algunos lugares de los Estados Unidos deben aprobar un test de habilidades básicas (Pre-Professional Skills Test, PPST) de las materias a enseñar (National Teacher Examination, NTE) y de conocimientos pedagógicos (batería básica del NTE). Pero la investigación ha mostrado que tener buenos resultados en esas pruebas tiene poca relación con la calidad del trabajo docente detectada con otros medios, incluyendo resultados de los alumnos. Esto no sorprende, pues las pruebas tienen principalmente preguntas de respuesta estructurada, como las de opción múltiple; los cuestionamientos que se hacen a esas pruebas cuando se trata de valorar el rendimiento de los alumnos se aplican con más razón a las que pretenden evaluar la calidad de los docentes.

Para mejorar los protocolos de observación usados por directores e inspectores, se han desarrollado también *evaluaciones de desempeño*, tanto con instrumentos de *baja inferencia*, con los que sólo se registra la frecuencia de unas conductas, como de *alta inferencia*, con los que el observador evalúa las conductas a medida que observa y registra su apreciación al respecto (cfr. Rosenshine y Furst, 1973).

Ejemplos de protocolos de observación de este tipo son el Florida Performance Measurement System (FPMS), los Teacher Performance Appraisal Instruments de Georgia y el Appraisal System de Texas.

Aunque mejores que los instrumentos rudimentarios utilizados en el pasado, las listas de conductas cuya presencia o ausencia se busca verificar mediante esos instrumentos siguen centrando la atención en aspectos superficiales de la docencia y suponen un solo modelo de buenas prácticas.

Para resultar bien calificados con los Teacher Performance Appraisal Instruments, los maestros de Georgia deben mostrar simplemente que pueden implantar un currículo prescrito externamente, y su habilidad para hacer cosas como "comenzar la clase a tiempo", "mantener un ritmo dinámico de instrucción", "manejar rutinas" y "redactar objetivos conductuales" (Darling-Hammond, 2008: 757).

El sistema de observación de Florida comenzó aplicándose en 1982 a maestros de nuevo ingreso a la profesión, y luego se extendió a la observación de maestros veteranos. Darling-Hammond y Sclan (1992: 20) señalan que, aunque el desarrollo del sistema partió de una exhaustiva revisión de la investigación sobre docencia, el FPMS deja fuera aspectos relevantes de la práctica por su dificultad para ser observados, como el conocimiento que tiene el docente del contenido; la planeación; el tipo de tareas asignadas y de la retroalimentación ofrecida; la calidad y variedad de materiales utilizados; la atención de los alumnos con dificultades, y el uso de estrategias para estimular actividades que impliquen un nivel cognitivo alto, entre otros.

La misma autora ofrece ejemplos de la manera en que el manual de codificación del FPMS orienta al observador para que registre como inadecuadas acciones que el documento mismo señala que pueden ser adecuadas, desalentando que la enseñanza se diversifique según las circunstancias de los alumnos, y orientando al docente hacia una práctica mecánica y estereotipada, en dirección opuesta a la que apunta la mejor investigación. Un indicador claro de las fallas del FPMS fue el contraste entre sus resultados y otras valoraciones, como ocurrió en el caso de un docente que recibió el premio de Maestro del Año de Florida en 1986, y fue calificado negativamente con ese instrumento (Darling-Hammond y Sclan, 1992: 19-21).

El uso de este tipo de protocolos de observación para sustentar decisiones de alto impacto sobre los maestros ha sido cuestionado y ha hecho que dejen de usarse en ese sentido, a veces como resultado de decisiones judiciales, como ocurrió con el sistema de evaluación de Georgia, cuya base fue considerada insuficiente por la corte del Condado Fulton, que lo encontró "arbitrario y caprichoso", al resolver una controversia al respecto (Kitchens vs. State Dept. of Education, Civil Action File # D-54773, 29 de julio de 1988, en Darling-Hammond y Sclan, 1992: 22, nota 5).

Esquemas de pago por méritos

La idea de que los maestros no deberían recibir una remuneración igual, sino una diferenciada, que reconozca el mérito de cada uno, no es una invención reciente. Donald Gratz (2009: 47-49) señala que, desde principios del siglo XIX, en algunos lugares de Inglaterra la paga de los maestros se basaba, en parte, en los resultados de sus alumnos en exámenes de lectura, escritura y aritmética. Esa práctica siguió en vigor durante mucho tiempo y quedó plasmada en una importante norma: el Revised Education Code, de 1862. Cinco años después a las áreas tradicionales se añadieron la historia, la geografía y la gramática. Según el mismo autor, siguiendo a Wilms y Chapleau (1999):

Las escuelas recibían una cantidad por cada niño que era evaluado satisfactoriamente en los exámenes, y eran penalizadas por tener tasas de asistencia bajas o resultados insatisfactorios. En el nivel elemental, por el desempeño "satisfactorio" de cada niño la escuela recibía seis chelines y seis peniques; el premio por cada niño de más edad era de 12 chelines [...] la penalización por el desempeño insatisfactorio de un alumno en lectura, escritura y aritmética era de dos chelines ocho peniques por materia (Gratz, 2009: 49).

Gratz añade que el sistema inglés no ofrecía actualización a los maestros, muchos de los cuales tenían mala preparación, por lo que la estrategia para tener mejores resultados consistió en aplicar mayor presión sobre los alumnos y en una práctica intensiva para que se prepararan para "el temido día de los exámenes", con consecuencias negativas que se fueron agravando con el tiempo (2009: 49).

Según un inspector, "el Código pensaba por el maestro, diciéndole en detalle lo que debía hacer", de manera que los maestros tendían a eliminar la enseñanza del dibujo, las ciencias o el canto, dedicando toda su atención a las áreas del examen, de la manera más mecánica, ya que según los maestros la memorización y la repetición daban los mejores resultados.

A medida que la presión sobre alumnos y docentes se agravó, como resultado del Código Revisado, se produjo una reacción igualmente creciente por parte de los maestros que, con otros factores, llevó a la creación del Sindicato Nacional de Maestros (National Union of Teachers) en 1870. La oposición del Sindicato, con la disminución del apoyo que inicialmente tuvo el sistema de exámenes por parte de la población a medida que la educación se volvió efectivamente universal, llevó finalmente a la supresión del sistema en 1892 (Gratz, 2009: 50-51).

A fines del siglo XIX en Canadá también se puso en práctica un sistema de pago a los maestros por los resultados de sus alumnos, que

produjo un aumento substancial en los resultados en los exámenes (en buena parte porque los maestros concentraban sus esfuerzos en los alumnos que tenían más probabilidades de tener éxito) junto con una reducción igualmente substancial del currículo. También en este caso el experimento fue abandonado como resultado de un fuerte clamor popular en ese sentido (Gratz, 2009: 52).

En los Estados Unidos la forma básica de definir el salario de los maestros durante el siglo XX fue la escala única, pero hubo numerosos intentos de combinarla con diferentes formas de pago por méritos.

De hecho se puede hablar de una oscilación permanente (*swinging pendulum*) entre los dos sistemas que refleja la idea de que lo deseable sería un sistema de pago que reconociera de manera indudable los méritos individuales, pero como no se encuentra una forma de detectar de manera perfectamente confiable esos méritos, se regresa a la escala salarial única, como un sistema más realista y viable, aunque no perfecto (*cfr.* Stronge, Gareis y Little, 2006: 8-10).

La primera experiencia de pago por mérito durante el siglo XX se registra, según Fenwick English (1992) en Newton, Massachusetts, en 1908. En la década de 1960 alrededor de 10% de los distritos escolares del país desarrollaron alguna variante de pago por méritos, pero en 1972 la proporción había bajado a 5.5%, y en 1979, a 4% (Gratz, 2009: 62).

Tras la publicación del informe "Una nación en peligro" (*A Nation at Risk*, 1983), y aprovechando protocolos de observación como los descritos en el inciso anterior, el interés por sistemas de pago por mérito se reavivó en varios estados, al pensar que ya se podía contar con instrumentos para identificar de manera confiable a los maestros ejemplares.

Como había ocurrido antes, los sistemas de pago por mérito de la década de 1980 tuvieron también una existencia precaria, y la mayor parte dejó de funcionar en un tiempo relativamente corto. La duración promedio de 183 casos analizados en 1978 y 1979 por el Educational Research Service era de sólo seis años. En el año 2000 Odden decía que "era difícil encontrar muchos ejemplos de cambios en la forma de pagar a los maestros implantados antes de 1990 que sobrevivan hoy" (según Gratz, 2009: 63).

La idea de diferenciar el salario del docente con base en la calidad de su trabajo mantiene su atractivo, pero sigue enfrentando el problema de distinguir de manera confiable a los maestros buenos y a los menos buenos. La aplicación de pruebas de rendimiento a todos los alumnos de ciertos grados que se ha extendido en los Estados Unidos y otros países a partir de la última década del siglo XX ofrece nuevos medios para tal efecto, como se verá en seguida.

En países que tienen un menor nivel de desarrollo, y en los que los sistemas educativos enfrentan todavía el reto de la cobertura universal en el nivel de primaria, los esquemas de pagos diferenciales se han utilizado para estimular que los maestros acepten trabajar en escuelas situadas en zonas muy pobres o remotas (Bolivia), o para reducir las inasistencias de los docentes (India, Kenia, Nicaragua) (*cfr.* Santibáñez, 2010; Morduchowicz, 2011).

La idea del pago por méritos para estimular el desempeño de los maestros de educación básica considera aplicables en el campo educativo estrategias que han dado resultados en otros ámbitos, y parece atractiva intuitivamente.

Además de que las experiencias reales de pago por mérito no parecen haber dado los resultados esperados, sus críticos apuntan razones de principio que hacen dudar de su efectividad: que el efecto de un estímulo material depende de algunas condiciones que no necesariamente se cumplen en educación, como que los

maestros tengan la capacidad de trabajar mejor, y no lo hagan simplemente por minimizar el esfuerzo; que hay bases para considerar que el estímulo más poderoso en el caso de la docencia es intrínseco, y que los estímulos individuales pueden tener consecuencias negativas, al debilitar la disposición para trabajar en equipo, que parece fundamental en las escuelas.

A estas razones se añade la dificultad ya mencionada para distinguir de manera confiable a los mejores maestros, con el riesgo de que los estímulos se entreguen a personas que no son las que más los merecen.

Por lo anterior hay una tendencia a tener cautela para utilizar pago por mérito y, en particular, para manejar esquemas de estímulos para los colectivos escolares, y no para los maestros en lo individual (cfr. Morduchowicz, 2011).

Sistemas de indicadores múltiples

Las limitaciones de los sistemas de pago por mérito y, en general, de los esfuerzos anteriores para detectar a los mejores maestros, pueden haber sido catalizadores de los replanteamientos sobre evaluación de docentes desarrollados a partir de mediados de la década de 1980 (Ladson-Billings, 2009: 215).

El año de 1986 representó el inicio de innovaciones más amplias por el impulso de dos grupos de trabajo: el llamado Holmes Group y, sobre todo, uno patrocinado por el Foro Carnegie sobre Educación y Economía (Task Force on Teaching as a Profession), inspirado sobre todo en el trabajo de Lee Shulman y sus colaboradores en la Universidad de Stanford, el Teacher Assessment Project (Shulman y Sykes, 1986; Shulman, 1987a).

De esos trabajos surgieron sistemas de evaluación complejos que combinan pruebas de conocimientos básicos, de los contenidos a enseñar y pedagógicas con protocolos de observación más completos y con portafolios de evidencias del trabajo del evaluando, como planes de clase, cuadernos de alumnos, registros de clases, etcétera. Se puede mencionar el sistema Praxis, desarrollado por el Educational Testing Service; el Interstate New Teacher Assessment and Support Consortium (INTASC), iniciado por California y Connecticut, seguidos por otros estados americanos, y el sistema del National Board for Professional Teaching Standards (NBPTS), desarrollado desde 1987 a partir del trabajo de Shulman y sus colaboradores en la Universidad de Stanford, con el apoyo inicial de la Fundación Carnegie y de varias organizaciones de maestros, y luego con el del gobierno federal (NBPTS, 1989; Ingvarson y Hattie, 2008).

El punto de partida de toda evaluación es una definición precisa del referente con el que se contrastarán las mediciones observadas empíricamente. En el caso de la evaluación de docentes, la definición de lo que es un buen docente, de los rasgos que conforman la competencia docente, es ese punto de arranque; sólo después habrá que buscar las mejores formas de obtener información sobre el grado en que el trabajo de ciertos maestros reúne o no esos elementos.

Por ello el desarrollo de sistemas más completos de evaluación parte de precisar la imagen del buen maestro que será su referente esencial. Esas definiciones toman la forma de estándares de las buenas prácticas docentes; cada sistema de evaluación tiene su propia versión al respecto, si bien hay un considerable grado

de coincidencia. La versión de los estándares del NBPTS se sintetiza en cinco puntos que definen los rasgos que los buenos maestros deben reunir:

- Estar comprometidos con los alumnos y su aprendizaje.
- Conocer las materias que enseñan y cómo enseñarlas a sus alumnos.
- Ser responsables de conducir y monitorear el aprendizaje de sus alumnos.
- Reflexionar sobre su propia práctica y aprender de la experiencia.
- Ser miembros de comunidades de aprendizaje (NBPTS, 1989).

Cada uno de estos rasgos generales se precisa con diversos grados de detalle, hasta llegar a definiciones operacionales muy precisas que permiten desarrollar instrumentos de obtención de información al respecto.

Una versión todavía sintética, pero que desarrolla los cinco enunciados anteriores en una forma que permite apreciar mejor el sentido del referente del sistema de evaluación del NBPTS, es la siguiente:

- **Los buenos maestros están comprometidos con los alumnos y su aprendizaje.** Los maestros certificados por el NBPTS están dedicados a hacer accesible el conocimiento a sus alumnos, a los que tratan de manera equitativa reconociendo sus diferencias. Adaptan su práctica docente con base en su observación y conocimiento de los intereses, habilidades, competencias, conocimientos, circunstancias familiares e interrelaciones de sus alumnos. Entienden cómo aprenden y se desarrollan éstos. Son conscientes de la influencia del contexto y la cultura en las conductas de sus estudiantes, cuya capacidad cognitiva y respeto por el aprendizaje desarrollan. Con igual importancia, promueven la autoestima, la motivación, el carácter, la responsabilidad cívica y el respeto de sus alumnos por las diferencias individuales, culturales, religiosas y raciales.
- **Los buenos maestros conocen las materias que enseñan y cómo enseñarlas a sus alumnos.** Los maestros certificados por el NBPTS tienen una comprensión rica de la(s) materia(s) que enseñan, y pueden apreciar cómo se crea el conocimiento de esas materias, cómo se organiza y se relaciona con otras disciplinas, y cómo se aplica en circunstancias reales. Dominan los conocimientos especializados y cómo transmitirlos o revelarlos a los alumnos. Tienen conciencia de los preconceptos y los conocimientos generales subyacentes que los alumnos típicamente tienen previamente, así como de los materiales y estrategias que pueden servir de apoyo para la enseñanza. Su repertorio de estrategias de enseñanza les permite crear diversos acercamientos al conocimiento, y siempre buscan enseñar a los alumnos cómo plantear y resolver sus propios problemas.
- **Los buenos maestros son responsables de la conducción y el monitoreo del aprendizaje de sus alumnos.** Los maestros certificados por el NBPTS crean ambientes que atraen y mantienen el interés del alumno y permiten usar el tiempo de la manera más efectiva. Dominan una amplia gama de técnicas de enseñanza, saben cuándo es apropiada cada una, y pueden ponerlas en práctica cuando se requiere. Saben motivar a sus alumnos e involucrarlos individual y grupalmente asegurando un ambiente constructivo de aprendizaje;

saben también cómo organizar la enseñanza para permitir alcanzar las metas que la escuela establece para sus alumnos. Saben cómo motivar a los alumnos para aprender y cómo mantener su interés incluso ante fracasos temporales. Evalúan regularmente el progreso de cada alumno, así como el del conjunto del grupo. Utilizan diversos métodos para medir el aprendizaje y el progreso de los alumnos, y pueden explicar con claridad el desempeño de los estudiantes a sus padres.

- **Los buenos maestros reflexionan sistemáticamente sobre su práctica y aprenden de su experiencia.** Los maestros certificados por el NBPTS son ejemplo de virtudes que buscan inspirar en sus alumnos —curiosidad, tolerancia, honestidad, justicia, respeto por la diversidad y valoración de las diferencias culturales— así como de cualidades que son prerequisites del crecimiento intelectual: capacidad de razonar y de adoptar perspectivas varias, ser creativos y tomar riesgos, así como adoptar una orientación experimental y de resolución de problemas. Buscando siempre fortalecer su enseñanza, examinan críticamente su propia práctica, buscan consejo de otros, y aprovechan los resultados de la investigación educativa para ampliar su repertorio de recursos, profundizar sus conocimientos, afinar sus juicios y adaptar su enseñanza a los nuevos hallazgos, ideas y teorías.
- **Los buenos maestros son miembros de comunidades de aprendizaje.** Los maestros certificados por el NBPTS contribuyen a la eficacia de su escuela trabajando en forma colaborativa entre sí y con otros profesionales sobre políticas educativas, currículo y desarrollo profesional. Pueden valorar el avance de la escuela y la asignación de recursos escolares a la luz de su comprensión de los objetivos educativos nacionales y locales. Tienen conocimiento de los recursos especializados de la escuela o la comunidad que se pueden aprovechar en beneficio de los alumnos, y tienen la capacidad de utilizar esos recursos según sea necesario. Encuentran formas de trabajar colaborativa y creativamente con los padres de familia, involucrándolos productivamente en el trabajo escolar (NBPTS, 1989).

Es evidente la diferencia que hay entre las dimensiones de las escalas de salarios tradicionales y los elementos de la definición del NBPTS, que también está lejos de la enumeración de rasgos del buen maestro, a los que reduce al conocimiento de las materias y de principios pedagógicos. La lista de rasgos del NBPTS y otras que subyacen los sistemas de evaluación más completos que hoy se manejan incluyen desde luego esos rasgos, pero también muchos más, como se acaba de ver.

A partir de los rasgos generales anteriores, en un nivel de concreción más preciso el NBPTS ha desarrollado más de 30 conjuntos de estándares específicos, según distintas áreas de contenidos curriculares y niveles de edad, desde preescolar hasta el final de la secundaria superior.

El sistema del NBPTS pretende identificar a los docentes más calificados, con propósitos de certificación avanzada; para ello, a diferencia de sistemas anteriores de evaluación que manejaban concepciones más simples de lo que es una buena

enseñanza y la reducían a la implantación de conductas básicas, propone una visión que destaca la importancia de emplear una amplia base de conocimientos de manera estratégica en el aula, lo que implica la presencia de juicios profesionales aplicados a casos particulares.

Los instrumentos para determinar si las prácticas de un maestro satisfacen los estándares del NBPTS incluyen viñetas con descripciones precisas de acciones de maestros, evidencias de la práctica reunidas en un portafolios que refleje el trabajo en aula a lo largo de varios meses —el cual debe incluir grabaciones en video, muestras del trabajo de los alumnos y entradas con reflexiones del evaluando—, así como evaluaciones de desempeño que se llevan a cabo en centros autorizados para ello (Darling-Hammond, 2008: 765).

Los instrumentos del NBPTS han sido desarrollados, piloteados y validados a lo largo de dos décadas mediante un extenso trabajo de investigación en el que han participado muchos de los investigadores más destacados del campo, y que ha sido reportado en numerosas publicaciones (*cfr.* Ingvarson y Hattie, 2008).

Aunque obviamente no es perfecto, opiniones calificadas consideran que el sistema del NBPTS “posiblemente sea el mejor sistema de evaluación del desempeño docente jamás desarrollado” (Pacheco, 2009: 166). En la introducción de la obra citada, que reporta ampliamente el desarrollo del sistema, los editores destacan la importancia de la actuación del primer presidente del National Board, Jim Kelly, que fue capaz

de convencer a las agencias que financiaron el trabajo, incluyendo a gobiernos federales demócratas y republicanos de los Estados Unidos [...] de que fueran pacientes; de que el trabajo sólo valdría la pena si se hacía bien [teniendo en cuenta que] establecer plenamente un esquema nacional de certificación llevaría 20 años (Ingvarson y Hattie, 2008: 3).

El Interstate New Teacher Assessment and Support Consortium comenzó también en 1987, y desarrolló estándares para la práctica profesional del maestro, y un sistema de evaluación que incluye un portafolios.

En 1988 el Educational Testing Service (ETS) inició a su vez el desarrollo de un sistema que incluye versiones enriquecidas de sus pruebas para evaluar niveles básicos de lectura, escritura y matemáticas, de áreas del currículo y de conocimientos pedagógicos, con preguntas de opción múltiple, pero también de respuesta abierta. El sistema Praxis III incluye entrevistas y observaciones en aula, así como la preparación de un portafolios por parte de cada evaluando.

La principal diferencia entre el NBPTS, el INTASC y el Praxis III radica en que el primero de estos tres sistemas pretende reconocer a los maestros sobresalientes, en tanto que los otros dos tienen como destinatarios maestros de menor experiencia y desempeño normal, para asegurar un mínimo de competencias con los fines tanto de dar la licencia para el ejercicio profesional como de aceptar a los candidatos que se incorporan a la profesión.

Modelos de valor agregado

Hasta hace pocos años ningún sistema educativo grande contaba con información de buena calidad para sustentar sistemas de remuneración de los docentes que no se basaran en indicadores como los de las *cuatro eses* de Mancur Olson.

La aplicación de pruebas estandarizadas de rendimiento a grandes muestras de alumnos, o incluso a todos los de ciertos grados, que se ha extendido en muchos países desde la década de 1990, ha llevado a proponer una forma alternativa de valorar el desempeño de los maestros, sin tener que observar a cada uno en su salón de clases ni basarse sólo en escolaridad y antigüedad.

La idea básica es sencilla y a primera vista indiscutible: el fin último del trabajo docente es que los alumnos alcancen ciertos aprendizajes; por tanto, si se cuenta con información comparable y confiable sobre el rendimiento de los alumnos, no debería ser necesario observar el trabajo de cada maestro para valorar su propio desempeño: bastaría inferirlo a partir del rendimiento de los alumnos. A la objeción de que algunos alumnos comienzan el curso en una posición mejor que otros, por lo que no es justo comparar los resultados de todos al final, se responde diciendo que lo que se debe comparar no son los resultados de cada alumno al término del curso, sino la diferencia entre lo que sabía al inicio y lo que sabe al fin del mismo, diferencia que es el *valor agregado* (*value added*) atribuible al trabajo del docente.

Pese al atractivo de la idea a primera vista, un análisis cuidadoso muestra que esos modelos tampoco dan información suficiente sobre el desempeño de los maestros, y menos sobre sus prácticas. Aunque hay diferencias en puntos menores, la afirmación anterior refleja el consenso actual de los especialistas en el sentido de que en la práctica, en el contexto real de los sistemas educativos actuales, no es factible implementar un sistema confiable de evaluación de docentes basado en modelos de valor agregado por varias razones:

- Limitaciones de las pruebas en gran escala que no miden aspectos importantes del rendimiento de los alumnos, incluyendo actitudes, pero también competencias cognitivas complejas.
- Imprecisión de las mediciones y su inestabilidad en el tiempo que hace que los márgenes de error de las estimaciones del valor supuestamente agregado por los maestros sean demasiado grandes para servir de base sólida para la toma de decisiones sobre individuos.
- Dificultad práctica para contar con bases de datos completas, en especial para seguir individualmente a los alumnos que cambian de escuela.
- Dificultad de principio para atribuir el avance de los alumnos en un grado a un solo maestro, dado que alumnos o maestros pueden cambiar, puede haber más de un maestro, y en los resultados de un grado cuenta la influencia de los maestros anteriores.

Un argumento más sobre la insuficiencia de los acercamientos de valor agregado es que, incluso suponiendo que la información que proporcionan fuera confiable para distinguir maestros más y menos eficaces, lo que bastaría para propósitos

de evaluación sumativa, no lo sería para propósitos de mejora, puesto que no permitiría saber en qué consisten las mejores prácticas para tratar de extenderlas.

Para ver argumentos en pro y en contra de los modelos de valor agregado pueden consultarse Rubin, Stuart y Zanutto (2004); Braun (2005); Glazerman *et al.* (2010); Leckie y Goldstein (2009) McCaffrey *et al.* (2003), y Rothstein (2011). En un reciente número de la revista *Phi Delta Kappan* que tiene como tema central el tema de la evaluación de maestros, Ritter y Shuls (2012) defienden el uso de modelos de valor agregado argumentando que, pese a sus indudables limitaciones, serían una alternativa mejor —o menos mala— que las otras que hay para dicho propósito, pero en el mismo número Paige (2012) destaca la seriedad de los problemas legales que enfrentarán las decisiones sobre maestros individuales que se tomen con base en resultados de esos modelos, y Linda Darling-Hammond y sus colaboradores (2012) reiteran los argumentos sustantivos que recomiendan el uso de los modelos de valor agregado con propósitos de investigación y validación de otros acercamientos a nivel grupal, pero no para decisiones individuales. ●

Capítulo 2

En busca de mejores sistemas de evaluación

En las primeras décadas del siglo XXI la preocupación por la calidad educativa ha seguido presente en los sistemas educativos, e incluso se ha acentuado, en parte probablemente por la difusión que se ha dado a los resultados de evaluaciones internacionales, como las de la OCDE (PISA) y la IEA (TIMSS y PIRLS).

Siguen presentes los dos grandes enfoques a la evaluación de los docentes cuyo desarrollo ha podido apreciarse en los apartados anteriores: el que se basa en el trabajo de cuerpos de inspectores, y el que utiliza instrumentos estandarizados de obtención de información sobre los maestros (sus conocimientos y prácticas) y los alumnos (su rendimiento). Ambos enfoques han tenido desarrollos, y hay señales de posibles combinaciones de ambos, como se mostrará a continuación.

EVOLUCIÓN DE LOS SISTEMAS DE INSPECTORES

En la Unión Europea los inspectorados siguen teniendo una presencia importante; la Standing International Conference of Inspectorates agrupa a 32 organismos de 23 países del viejo continente: Albania, Alemania (Hessen, Renania-Palatinado, Sajonia y Baja Sajonia), Austria, Bélgica (regiones de habla holandesa, francesa y alemana), Bulgaria, Dinamarca, España, Estonia, Francia, Irlanda, Italia (Trentino), Lituania, Malta, Noruega, Países Bajos, Portugal, Reino Unido (Inglaterra, Escocia, Gales e Irlanda del Norte), República Checa, República Eslovaca, Rumania, Serbia, Suecia y Suiza (SICI, 2013a).

Un estudio reciente (Van Bruggen, 2010) describe características y funciones de 18 de esos inspectorados: los de Dinamarca, España, Estonia, Irlanda, Noruega, los Países Bajos, Portugal, la República Checa, la República Eslovaca y Suecia, así como los de la región flamenca de Bélgica, los de los estados alemanes de Hessen, Renania-Palatinado y Sajonia, y los de Inglaterra, Escocia, Gales e Irlanda del Norte.

Llama la atención que únicamente el inspectorado de Inglaterra (Office for Standards in Education, Children Services and Skills, OFSTED) y, en menor medida, los de Escocia y Suecia no dependen legalmente del Ministerio de Educación,

pero casi todos gozan de independencia funcional, salvo los de Estonia e Irlanda del Norte (Van Bruggen, 2010: 134-138).

La misión de casi todos los sistemas se define como una de asesoría al gobierno, pero muchos también tienen asignada la tarea de retroalimentar a las escuelas, y en muchos casos se especifica que deben prestar especial atención a las que tienen bajos resultados u otros síntomas de debilidad, o en relación con las que han recibido quejas. En prácticamente todos los casos se difunden los resultados del trabajo (Van Bruggen, 2010: 139-142).

En general la observación de la práctica docente forma parte de la tarea de los inspectores, si bien la cantidad, la duración y la forma de hacer las observaciones varían mucho, al igual que la forma de informar a los maestros sobre los resultados de éstas (Van Bruggen, 2010: 130-133).

La tarea de los inspectores incluye siempre la evaluación de las escuelas como tales, para lo que equipos de inspectores (por lo general de 2 a 5, pero pueden llegar hasta 10) deben realizar visitas con diferente frecuencia y, por lo general, durante varios días, para revisar distintos aspectos del funcionamiento del plantel que están plasmados en un marco de referencia conocido. Las formas de dar retroalimentación a la escuela también varían, pero por lo general hay un informe que se hace público (Van Bruggen, 2010: 121-129).

En los Estados Unidos los sistemas de inspectores dejaron el lugar a modelos de evaluación diferentes, aunque en muchos casos existen instancias que cumplen funciones de asesoría y apoyo. Sólo hay sistemas de inspección como tales en Nueva York, Massachussets y Rhode Island. Sin embargo, las limitaciones de los esquemas más sofisticados de evaluación, a los que se hace referencia más abajo, están dando lugar a un renovado interés por los sistemas de inspectores, no como alternativa, pero sí como posibles componentes de un sistema más completo de evaluación. Es interesante advertir que el gobernador de California vetó una ley aprobada por el Senado para hacer aún más complejo un sistema de evaluación de maestros basado en modelos de valor agregado (Brown, 2011), después de que el Centro de Políticas Aplicadas en Educación de la Universidad de California en Davis organizó una reunión sobre los inspectorados, con el título *¿Una solución británica para las escuelas públicas de California?* (CAP-Ed, 2007).

INVESTIGACIONES SOBRE LA EFECTIVIDAD DE LOS DOCENTES

El camino norteamericano para evaluar la efectividad de los docentes ha sido, pues, diferente. En lugar de sistemas de inspectores ha partido de la investigación sobre los maestros, dando lugar al desarrollo de instrumentos estandarizados de obtención de información sobre la práctica docente, pero las herramientas más complejas, como los modelos de valor agregado, e incluso los sistemas basados en indicadores múltiples como el del National Board for Professional Teaching Standards (NBPTS), no son plenamente satisfactorios.

Los tres componentes básicos de los sistemas de indicadores múltiples (pruebas aplicadas a los maestros, protocolos de observación y análisis de evidencias

del trabajo de los evaluandos) presentan una debilidad común: la dificultad de captar los aspectos más finos de la práctica docente, en particular cuando se trata de desarrollar competencias complejas y no sólo de asegurar que los alumnos memoricen información o apliquen procedimientos de manera mecánica.

Como ocurre con las pruebas del rendimiento de los alumnos, las que se aplican a maestros para verificar que empleen los contenidos a enseñar y la pedagogía correspondiente también se pueden limitar a valorar la capacidad memorística y el conocimiento de datos irrelevantes para la docencia. Investigaciones recientes señalan que las pruebas usuales no logran captar una dimensión del conocimiento de los docentes que es básica para que la enseñanza sea mejor: el *conocimiento pedagógico de los contenidos*. Hace más de un cuarto de siglo Shulman advertía sobre la importancia de este elemento al comentar una expresión usual según la cual las personas que se dedican a la enseñanza lo harían por no ser capaces de hacer algo más productivo; este investigador replica que para enseñar unos contenidos no basta “conocerlos”, sino que es necesario “comprenderlos” en un grado suficiente para poder explicarlos a los estudiantes (Shulman, 1986).

La misma idea expresa la frase “conocimiento de las matemáticas para la enseñanza” (*mathematical knowledge for teaching*), que Ball y Hill explican así:

No es el conocimiento pedagógico en general; es, sin duda, conocimiento de contenido matemático, pero una forma específica que los maestros necesitan para poder explicar soluciones, detectar fallos típicos en la forma de razonar de un alumno y explicar conceptos clave [...] Incluye la habilidad de resolver problemas matemáticos que se presentan a los alumnos o a las personas adultas, pero no se reduce a ella, y tampoco es igual a la capacidad de manejar matemáticas más complejas (Lazer, 2009: 153).

Los autores de la expresión “conocimiento de las matemáticas para la enseñanza” ilustran su significado con ejemplos elocuentes:

el conocimiento necesario para enseñar a un niño a multiplicar implica más que ser capaz de hacer esos cálculos uno mismo. También es más que dominar las matemáticas que comprende el currículo de cuarto de primaria y un poco más, como quieren algunos. Es una mezcla bastante compleja de conocimientos, habilidades, procesos de razonamiento y otros hábitos mentales, como la capacidad de ver lo que implica la perspectiva de otra persona y entender lo que está haciendo, qué necesita un maestro, y no un matemático o un físico (Ball y Hill, 2009: 88).

[...] detectar sutiles diferencias al elegir los ejemplos a utilizar en una clase es una tarea que implica una compleja habilidad matemática crítica para el trabajo docente, y que muchas personas con buena formación matemática no tienen, incluyendo a muchos investigadores. La enseñanza requiere una peculiar manera de ver los contenidos, que no es común en los que utilizan las matemáticas para otros propósitos (Ball y Hill, 2008: 90).

Por lo que se refiere a los protocolos de observación, ya se ha señalado que los más usuales se limitan a verificar si el maestro realiza conductas precisas que se consideran representativas de buenas prácticas, pero que reflejan un solo modelo de buena enseñanza, y no permiten apreciar si esas prácticas se llevan a cabo de tal manera que promuevan el desarrollo de competencias cognitivas complejas.

Como ya se ha apuntado, si la enseñanza se orienta al desarrollo de habilidades de bajo nivel cognitivo, se pueden precisar conductas fáciles de observar que el docente debe realizar, pero si se trata de desarrollar competencias complejas, los docentes deben poner en práctica estrategias docentes de gran complejidad también, cuya observación es igualmente difícil.

Enseñar de manera que los alumnos alcancen altos niveles de comprensión, pensamiento y razonamiento es desafiante para el docente [...] Lo que determina el nivel de aprendizaje no es si los alumnos trabajan en grupos o si manipulan objetos, es más bien la naturaleza de las tareas a las que se ven expuestos y si se mantienen altos niveles de demanda cognitiva [...] Las pruebas estandarizadas no se diseñan para cubrir toda la gama de contenidos y habilidades que los alumnos deben aprender a lo largo de un año escolar [...] sino que se limitan a franjas relativamente estrechas de contenido [...] Las pruebas tienden a concentrarse también en habilidades de bajo nivel (Stein y Matsumura, 2008: 189, 196 y 199).

En la perspectiva que se tiene actualmente de la misión de un sistema educativo, el desarrollo de competencias cognitivas complejas y no cognitivas ("habilidades blandas") constituye la dimensión más importante de una buena enseñanza, y es, a la vez, lo más desafiante para los docentes y lo más difícil de captar, incluso por los mejores instrumentos disponibles.

Por ello la búsqueda de mejores formas de identificar al buen docente sigue, con proyectos de investigación que atienden cuestiones cada vez más finas y diseños más complejos. De los primeros años del siglo XXI datan un estado de la cuestión (Porter, Youngs y Odden, 2001) y una revisión crítica de posturas en boga con influencia en la política nacional al respecto (Darling-Hammond y Young, 2002).

Más recientemente destacan nuevos balances (Goe, Bell y Little, 2008; Robinson y Campbell, 2010; Strong, 2011; Correnti y Martínez, 2012; Kelly, 2012; Martínez Rizo, 2012b). En enero de 2013 se difundió el informe final del Proyecto MET (Measures of Effective Teaching), patrocinado por la Fundación Bill & Melinda Gates (Kane y Staiger, 2013).

A continuación se organizan los elementos que conforman el perfil de un buen docente en una forma que facilite analizar la gama de instrumentos de obtención de información que utiliza un sistema de evaluación según los aspectos del perfil que atiende cada instrumento.

El perfil del buen docente incluye: a) rasgos personales como concepciones, actitudes y valores, y b) los conocimientos generales, los particulares de las áreas curriculares a enseñar y los que se refieren al manejo de la pedagogía necesaria para enseñarlos. Estos dos elementos se refieren a insumos necesarios para que haya una buena enseñanza, pero no incluyen resultados de ésta ni los procesos

en los que consiste. Por ello un tercer elemento del perfil de un buen maestro, que aparece sólo recientemente, es c) el relativo a los resultados de la enseñanza, en el sentido de los aprendizajes que alcanzan los alumnos. Un cuarto elemento consiste en d) las prácticas docentes en que se concreta el proceso de enseñanza. Y un elemento más, en e) los juicios sobre la calidad de los docentes del maestro mismo, sus pares, el director de la escuela y los supervisores, pero también los alumnos mismos y, en ciertos aspectos, los padres de familia.

Los instrumentos de obtención de información sobre esos aspectos se pueden organizar en tres grupos, según se basen en respuestas a preguntas orales o escritas, en la observación de lo que pasa en el aula, o en el análisis de productos del trabajo del docente y sus alumnos. La tabla 1 visualiza el espacio definido por rasgos del buen docente e instrumentos de obtención de información. Los símbolos que aparecen en las casillas sugieren la mayor o la menor adecuación de cierto tipo de instrumentos para la obtención de información sobre ciertos rasgos.

Tabla 1 Rasgos del perfil del buen docente e instrumentos de obtención de información al respecto

Rasgos del perfil del buen docente	Tipos de instrumentos de obtención de información		
	Basados en interrogación	Basados en observación	Basados en análisis de materiales
Rasgos personales	++	-	+
Conocimientos	++	-	+
Aprendizaje alumnos	++	+	++
Prácticas docentes	+	++	+
Juicios de varios actores	++	-	-

Fuente: elaboración propia.

La siguiente presentación de instrumentos de los tres tipos mencionados ofrecerá al lector una vista panorámica de los esfuerzos en curso sobre el tema. Para información más amplia al respecto, véase Martínez Rizo (2012b).

Acercamientos basados en interrogación

En ciencias del hombre, el cuestionario y la entrevista, con sus múltiples variantes, son instrumentos muy conocidos de obtención de información, y hay una extensa literatura sobre lo que hay que cuidar para obtener respuestas de la mejor calidad posible mediante ellos.

En la tabla 1 se aprecia que, en principio, este tipo de instrumentos es bastante adecuado para obtener información sobre los cinco aspectos mencionados del perfil del buen docente, con excepción de las prácticas docentes mismas.

Cuestionarios y escalas de actitud son indicados para explorar las concepciones de los docentes, así como sus actitudes y valores.

Las pruebas, que son un tipo de cuestionario, son apropiadas para estudiar tanto los conocimientos que aprenden los alumnos como los que tienen los maestros, en lo relativo a conocimientos generales básicos, a las materias que enseñan y a la pedagogía correspondiente. Sin embargo, no sobra reiterar la idea de que las formas más usuales de pruebas, formadas por preguntas de respuesta estructurada, por lo general se limitan a medir niveles cognitivos elementales. El desarrollo de pruebas para valorar competencias complejas supone personal muy especializado y largos tiempos de desarrollo. Hay ya ejemplos de instrumentos que miden aprendizajes complejos, pero hacerlo en gran escala enfrenta el reto de diseñar y calificar preguntas de respuesta construida y de ejecución, pruebas matriciales o ejercicios de simulación como los que el acceso masivo a computadoras portátiles permite.

Los cuestionarios son también apropiados, en principio, para recoger juicios sobre los docentes que pueden formular ellos mismos (autoevaluación), sus pares, los directores y supervisores, los alumnos y los padres de familia. En este caso es importante tener en cuenta que cada actor puede ser informante calificado de algunos aspectos, pero no de todos. En el caso de maestros, directores y supervisores se deberá cuidar la tendencia a responder lo deseable en lugar de lo real, los temas delicados o amenazantes, y la posible presencia de sesgos sistemáticos favorables o desfavorables.

Si al estudiar concepciones, actitudes y valores de los docentes para formular las preguntas de los cuestionarios se utilizan nociones derivadas de ciertos planteamientos teóricos, además del riesgo de recibir respuestas socialmente deseables, es muy probable que no haya una comprensión uniforme de las mismas por parte de los maestros, lo que pone en duda la confiabilidad de las respuestas.

Si se pregunta a unos maestros, por ejemplo, si utilizan *trabajo colaborativo* o *evaluación formativa*, y responden afirmativamente, es difícil saber qué entendían exactamente por esos términos, y es probable que haya una diversidad de comprensiones al respecto, lo que invalida seriamente las conclusiones que se saquen con base en tales respuestas.

Una forma de enfrentar este problema consiste en usar una variante de preguntas que se suele designar con el término de *viñetas*, con las cuales, en lugar de preguntar en términos abstractos, se utilizan descripciones precisas de conductas concretas, contextualizadas, las cuales piden a los sujetos que responden que indiquen si sus propias formas de trabajar se aproximan más o menos a las descritas en la pregunta o la viñeta.

El uso de viñetas mejora la calidad de la información obtenida, pero la elaboración de este tipo de preguntas implica un tiempo mucho mayor que el necesario para formular preguntas simples y, pese al cuidado que se ponga, es posible que la misma descripción en términos conductuales de un concepto abstracto no sea comprendida igual por distintos sujetos. Stecher et al. (2006) pidió a algunos maestros que leyeran viñetas preparadas para una investigación y expresaran su forma de entenderlas "pensando en voz alta". Encontró que en algunos casos los maestros no entendían de inmediato lo que quería decir la posible acción, y que antes de escoger una respuesta debían reformular la idea en sus propias palabras.

Encontró también que, para escoger una de las opciones de respuesta que se ofrecían sobre la mejor forma de actuar ante una situación como la descrita en la viñeta, veían necesario complementar la información que se les daba en ella con datos sobre las características de los alumnos con los que se trabajaría, ya que las acciones a emprender dependerían en buena parte de ello.

Otra limitación de los cuestionarios simples para estudiar prácticas docentes en un lapso amplio de tiempo, como un ciclo escolar, es el que se refiere a la dificultad de recordar con precisión lo que se ha hecho en momentos que no acaban de transcurrir. Una forma de enfrentar este problema consiste en la aplicación reiterada de cuestionarios breves, en la forma de diarios o bitácoras.

En este caso se pide al maestro que informe sobre sus actividades en periodos de tiempo cortos, pero que lo haga no una sola vez, sino varias, y lo más cerca que sea posible del momento en que ocurrieron los hechos, por ejemplo con una periodicidad diaria a lo largo de un lapso relativamente breve. De esta manera es más probable que los recuerdos sean precisos y la confiabilidad de la información recopilada aumenta, ya que es menos probable que se distorsione la realidad si las actividades se reportan varias veces a lo largo del tiempo. Para que las conductas reportadas sean representativas de las que realiza el maestro es necesario, desde luego, que este tipo de instrumentos se utilice durante un número suficiente de ocasiones, por ejemplo, en las clases diarias. Según Rowan y Correnti (2009), para tener información suficiente sobre las prácticas de un maestro son necesarios reportes de unos 20 días al año.

Las tecnologías de la información y la comunicación (TIC) actuales pueden facilitar considerablemente el trabajo de los maestros para responder diarios o bitácoras en línea. Con una plataforma adecuada, esta forma de recabar la información permite enviarla rápidamente y hace innecesario que los investigadores deban capturarla posteriormente.

Unos estudios en los que el uso de diarios y bitácoras es muy apropiado son los que se refieren al currículo implantado efectivamente, en contraposición al planeado. No es raro que los maestros durante el ciclo escolar cubran sólo una parte de los contenidos comprendidos en el programa de estudios respectivo por razones más o menos justificadas, como el exceso de contenidos, la escasa relevancia de algunos temas o lo inadecuado de otros para ciertos alumnos. Y es esperable que los alumnos no aprendan los contenidos no cubiertos. Por ello la distancia entre el currículo planeado y el implementado es una variable importante cuando se trata de explicar los resultados alcanzados por los estudiantes.

Un trabajo de Rowan, Camburn y Correnti (2004) ejemplifica el uso de bitácoras para el estudio del currículo implantado. Otro acercamiento a este tema es el de la Encuesta sobre el Currículo Implementado (Survey of Enacted Curriculum, Council of Chief State School Officers) (Blank, Porter y Smithson, 2001).

La principal aportación de esta encuesta radica en que no se limita a verificar si cierto tema incluido en el programa de estudios fue cubierto efectivamente o no por un maestro, sino que explora la profundidad con la que se hizo. El punto sobre el que el estudio llama la atención es que la diferencia entre lo planeado y

lo que se implanta efectivamente en el aula puede radicar no en el que un tema se cubra o no, sino en la manera en que se trate.

Los currículos actuales suelen manejar las nociones de competencias o de estándares buscando evitar limitaciones usuales en el pasado, cuando sólo se esperaba que los estudiantes fueran capaces de memorizar datos puntuales o de realizar ciertas operaciones en forma mecánica. Los programas contemporáneos plantean generalmente propósitos más ambiciosos, en el sentido de que los alumnos alcancen niveles cognitivos altos y desarrollen competencias complejas.

Para alcanzar esas metas ambiciosas, sobre todo en contextos vulnerables, los maestros no se pueden limitar a desarrollar los temas mediante exposiciones frontales u otras estrategias tradicionales limitadas. Es necesario que pongan en juego una extensa gama de técnicas, y en particular, que lo hagan de manera variada, en función de las necesidades, las posibilidades y los avances reales de sus alumnos. Por ello el análisis de las prácticas docentes en la Encuesta sobre el Currículo Implementado se hace con base en un espacio conceptual bidimensional en el que cada contenido se combina con el nivel de complejidad cognitiva esperado para el aprendizaje que alcancen los estudiantes.

Los resultados permiten apreciar cómo en muchos casos un tema se cubre, pero de una manera superficial, sin alcanzar los niveles planteados en el currículo. Éste es un buen ejemplo de acercamiento que permite información más rica que la que se suele obtener con los instrumentos convencionales.

Hay que añadir, sin embargo, que trabajos relativamente antiguos muestran que es posible obtener información de buena calidad mediante cuestionarios y autorreportes elaborados por los docentes sobre sus propias prácticas, si se cuidan los elementos que aumentan el riesgo de sesgos por autocomplacencia o tendencia a dar respuestas deseables (Mayer, 1999; Koziol y Burns, 1986).

Por otra parte, además de los maestros, otros sujetos pueden dar información de buena calidad sobre algunos aspectos de las prácticas docentes, como directores y supervisores, e incluso los alumnos mismos, sobre los temas cubiertos, las estrategias docentes utilizadas o la retroalimentación que les dan los maestros. En educación superior la aplicación de encuestas para conocer la opinión de los estudiantes sobre sus maestros tiene más de 70 años de historia, pero también en niveles educativos inferiores los alumnos pueden ser informantes valiosos, así como pueden serlo los padres sobre aspectos del hogar relacionados con la escuela.

Un ejemplo de instrumentos para obtener información de varias fuentes es el conjunto de cuestionarios del Proyecto Trípode para la Mejora Escolar (Tripod Project for School Improvement), desarrollado por el doctor Ronald F. Ferguson, de la Universidad de Harvard, para un consorcio de cientos de escuelas de distritos escolares de diversas regiones de los Estados Unidos (Ferguson, 2008).

El nombre de *trípode* se deriva del marco de referencia del trabajo, cuyo objetivo es conseguir que todos los alumnos tengan éxito en la escuela, y considera que eso depende de tres factores igualmente importantes, "tres patas" indispensables para que la tarea educativa se mantenga en pie:

- Su contenido: ¿qué hay que enseñar? ¿Tiene el maestro un conocimiento profundo de los contenidos curriculares?
- La pedagogía utilizada: ¿Cómo hay que enseñar? ¿Utiliza el maestro técnicas efectivas de instrucción?
- Las relaciones sociales en que se enmarca la actividad escolar: ¿somos una comunidad? ¿Se preocupan unos por otros, se inspiran y se motivan los maestros y los alumnos?

Sin descartar la importancia de otras técnicas como la observación, el Proyecto Trípode busca aprovechar otras fuentes de información: los alumnos, sus padres y los docentes mismos, una tríada de actores que pueden informar sobre las prácticas docentes. Se subraya la importancia del punto de vista de los alumnos, cuya vivencia cotidiana de las prácticas de sus maestros los hace informantes privilegiados, aunque no tengan la preparación de un observador profesional.

Los cuestionarios que se aplican a estudiantes de preescolar a secundaria superior indagan en especial sobre las condiciones del trabajo en el aula. Los que se aplican a los padres de familia centran la atención en las prácticas de crianza y en las relaciones entre el hogar y la escuela. Y los dirigidos a los docentes se refieren a las estrategias pedagógicas que utilizan, a las condiciones de su trabajo y al tipo de relación que tienen con sus alumnos (MET Project, 2010h).

Los cuestionarios del Proyecto Trípode se han afinado con base en los resultados de su aplicación a más de 400 000 estudiantes, a lo largo de más de 10 años. A partir de ello, sus autores consideran que hay bases sólidas para utilizar la opinión de los alumnos sobre el trabajo de sus maestros como medida de la efectividad de éstos (Ferguson, 2012).

Las prácticas de los docentes que se asocian a ciertas percepciones de los estudiantes se sintetizan en siete categorías que se expresan por palabras que en inglés comienzan con la letra c, por lo que se habla de las Siete C, como sigue:

Tabla 2 Las Siete C de prácticas docentes y experiencias del alumno

	¿Qué hace el maestro?	¿Qué experimenta el alumno?
Caring	Cuida a los alumnos	Que recibe aliento y apoyo
Controlling	Controla el comportamiento	Que necesita buena relación con amigos
Clarifying	Clarifica las lecciones	Que el éxito parece accesible
Challenging	Plantea retos	Que necesita esfuerzo y perseverancia
Captivating	Cautiva a los alumnos	Que las clases son interesantes
Conferring	Intercambia ideas con ellos	Que sus ideas son respetadas
Consolidating	Consolida el conocimiento	Que unos temas se conectan con otros

Fuente: Ferguson y Ramsdell (2011).

Acercamientos basados en observación

El segundo grupo de acercamientos al estudio de la práctica docente no se basa en lo que digan los maestros mismos u otros informantes, sino en la observación por terceras personas de dichas prácticas (*classroom observation*).

Desde el siglo XIX los principales sistemas educativos pusieron en marcha sistemas de inspección o supervisión para uniformar sistemas de enseñanza y asegurar mínimos de calidad. Una tarea habitual de los inspectores o supervisores era evaluar a los maestros, para lo cual se entrevistaban con cada uno, revisaban cuadernos de alumnos, planes de trabajo y otros materiales, etcétera.

Los supervisores visitaban los salones para observar la forma de trabajar de los docentes y verificar el grado de avance de los alumnos, mediante preguntas o pruebas escritas que ellos mismos desarrollaban. Usaban formas en que se ubicaba físicamente a los alumnos, como ayudas para la observación de lo que pasaba en el aula (*seating charts*). En 1914 Horn propuso usar símbolos para indicar si un alumno respondía una pregunta o realizaba cierta actividad. Las primeras formas que se desarrollaron para la observación sistemática de lo que pasaba en las aulas se basan en esas prácticas (*cfr.* Medley y Mitzel, 1963).

Hace medio siglo esos autores decían ya que “la investigación sobre las conductas que tienen lugar en el aula no es un pasatiempo para aficionados, sino una ocupación de tiempo completo para profesionales técnicamente competentes” (Medley y Mitzel, 1963: 253), y, en forma congruente, definían la observación como una técnica rigurosa en los términos siguientes:

Una técnica observacional [...] es una en la que un observador registra aspectos relevantes de una conducta que tiene lugar en el aula a medida que ocurre o muy poco después (*or within a negligible time limit after*) con un mínimo de cuantificación entre la observación de la conducta y su registro [...] por consiguiente no se incluyen en esta definición esquemas en los que se pide al observador que califique al maestro, a los alumnos o a la clase en una o más dimensiones, aun si las calificaciones se basan en la observación directa de conductas específicas (Medley y Mitzel, 1963: 253).

Con una concepción más amplia, la definición de observación podría incluir procedimientos con escalas de calificación; técnicas de registro y poscodificación (por ejemplo con grabaciones en video), y también técnicas cualitativas. Medley y Mitzel distinguían los sistemas de observación basados en tiempos (de categorías) y los basados en eventos (de signos).

Una década más tarde Rosenshine y Furst (1973) distinguían técnicas de observación según el procedimiento de registro: sistemas de conteo (categorías o signos) y de calificación (*rating*), según la especificidad de los ítems: sobre conductas específicas (de baja inferencia) o más generales (de alta inferencia) y, según el tipo de codificación, de una o más dimensiones.

Los instrumentos desarrollados recientemente suelen basarse en teorías actuales sobre los aspectos a observar, y cuidan asimismo las propiedades psicométricas. Su número es grande, pero en muchos casos han sido desarrollados

para una investigación particular: no se han usado en contextos variados y hay poca información que permita valorar su calidad. Por ello la descripción siguiente se reduce a algunos de los instrumentos más estudiados, seleccionados para el Proyecto sobre Medición de la Efectividad de la Enseñanza (MET Project, 2010a y 2010c), un ambicioso trabajo en proceso de desarrollo con participación de instituciones e investigadores destacados y financiamiento de la Fundación Bill & Melinda Gates.

Classroom Assessment Scoring System (CLASS)

Posiblemente la herramienta sobre la que hay más estudios. Es un instrumento para estudiar las actividades del docente y las interacciones que tiene con sus alumnos. Fruto de un trabajo de más de una década, a cargo de Robert Pianta y sus colegas, una versión previa se difundió con el nombre de Classroom Observation System —COS— (Pianta y Hamre, 2009; MET Project, 2010d).

La conceptualización de las actividades e interacciones distingue tres dominios: uno relativo a la organización del aula; otro, al apoyo instruccional que se ofrece a los alumnos, y uno más al apoyo emocional. La tabla 3 las sintetiza.

Tabla 3 Dimensiones de la enseñanza efectiva evaluadas por el CLASS

	Apoyo emocional	Organización del aula	Apoyo instruccional
Preescolar y primeros grados de primaria	<ul style="list-style-type: none"> ▪ Clima positivo ▪ Clima negativo ▪ Sensibilidad del maestro ▪ Atención al punto de vista del niño 	<ul style="list-style-type: none"> ▪ Gestión de conducta ▪ Productividad instruccional ▪ Formatos de aprendizaje 	<ul style="list-style-type: none"> ▪ Desarrollo conceptual ▪ Calidad en la retroalimentación ▪ Modelamiento lenguaje
Últimos grados de primaria y nivel secundaria	<ul style="list-style-type: none"> ▪ Clima positivo ▪ Clima negativo ▪ Sensibilidad maestro ▪ Atención al punto de vista adolescente 	<ul style="list-style-type: none"> ▪ Gestión de conducta ▪ Productividad instruccional ▪ Formatos de aprendizaje 	<ul style="list-style-type: none"> ▪ Comprensión de contenidos ▪ Análisis-solución de problemas ▪ Calidad en la retroalimentación ▪ Diálogo instruccional

Fuente: MET Project (2010d: 3).

La observación se lleva a cabo en periodos de media hora; en cada uno se dedican 20 minutos a observar y tomar notas, y el tiempo restante a la calificación. Según los autores bastan cuatro ciclos de observación para contar con una muestra representativa de lo que ocurre en un salón de clases.

El sistema se diseñó para observar grupos de preescolar o de los primeros grados de primaria; están en desarrollo versiones para grados superiores. La calidad de la información derivada del uso del sistema CLASS ha sido explorada empíricamente en forma extensa para la versión enfocada a preescolar y los primeros grados de primaria, con resultados satisfactorios. El trabajo respecto a las otras versiones está menos avanzado (Goe, Bell y Little, 2008: 24).

Marco de Referencia para la Enseñanza (Framework for Teaching)

A partir de su trabajo en el Educational Testing Service (ETS) para la Versión III del Praxis (*cf. supra*), Charlotte Danielson y sus colaboradores, autores de esta herramienta, la describen como: “estructurada según componentes de la instrucción derivados de la investigación, alineada a los estándares del consorcio interestatal (Interstate New Teacher Assessment and Support Consortium, INTASC), y basada en una perspectiva constructivista del aprendizaje y la enseñanza” (MET Project, 2010d; Goe, Bell y Little, 2008: 21-22).

El Marco tiene cuatro dominios: planeación y preparación de la clase; ambiente del aula; instrucción, y responsabilidades profesionales. Estos dominios se dividen en 22 componentes y 76 indicadores. A guisa de ejemplo, los componentes y los indicadores del dominio de ambiente del aula son los siguientes:

- Creación de un ambiente de respeto y buena relación
 - Interacciones del maestro con los alumnos
 - Interacciones de los alumnos entre sí
- Establecimiento de una cultura de aprendizaje
 - Importancia de los contenidos
 - Expectativas de aprendizaje y logro
 - Orgullo de los alumnos por su trabajo
- Procedimientos de gestión del aula
 - Manejo de grupos para la instrucción
 - Manejo de las transiciones entre etapas de los procesos
 - Manejo de materiales
- Gestión de la conducta de los alumnos
 - Expectativas
 - Monitoreo de la conducta de los alumnos
 - Respuesta ante las conductas inapropiadas (MET Project, 2010b)

El Marco incluye rúbricas detalladas con base en las cuales se puede evaluar a un maestro en cuanto a cada uno de los 76 elementos en 4 niveles de desempeño, que se definen como “insatisfactorio”, “básico”, “avanzado” (*proficient*) y “sobresaliente” (*distinguished*) (MET Project, 2010d).

Sistema MQI (Mathematical Quality of Instruction)

Como indica su nombre, la herramienta se orienta precisamente a la observación de prácticas de enseñanza en el campo de las Matemáticas, a partir de lo que la investigación al respecto ha encontrado (Hill *et al.*, 2010a y 2010b).

Se identificaron así cinco puntos como los más relevantes para caracterizar las prácticas docentes en este campo del conocimiento:

- uso preciso y rico del lenguaje matemático;
- ausencia de errores e imprecisiones;
- presencia de explicaciones matemáticas correctas y participación de los alumnos en la construcción de significado y en el razonamiento matemático;

- conexión del trabajo del aula con ideas importantes en el campo, y
- esfuerzos para garantizar que todos los alumnos del grupo, y no sólo algunos, tengan acceso al conocimiento matemático (MET Project, 2010e).

El MQI explora tres tipos de relaciones: de los maestros con los contenidos, de los alumnos con los contenidos, y de los maestros con los alumnos. Al valorar esas dimensiones, el MQI sobresale entre los instrumentos del área de Matemáticas, porque ofrece una visión completa y equilibrada de los elementos que, en conjunto, conforman la enseñanza de calidad en el área (MET Project, 2010e).

Durante el desarrollo del MQI se encontró que, además de un dominio general de las Matemáticas, el maestro debe tener un tipo particular de conocimiento matemático, precisamente *el que hace falta para la enseñanza de esta área*, que tiene que ver con una comprensión precisa de los obstáculos que dificultan a los alumnos el aprendizaje de este campo. Por ello en paralelo se ha desarrollado otra herramienta para medir ese tipo de conocimiento: el MKT (Mathematical Knowledge for Teaching) (MET Project, 2010e; Hill *et al.*, 2008 y 2004).

Sistema PLATO

Otro ejemplo de herramienta especializada en un área curricular particular es la denominada Protocolo para la Observación de la Enseñanza de las Artes del Lenguaje (Protocol for Language Arts Teaching Observation, PLATO).

Con base en investigaciones previas, el sistema PLATO se estructura a partir de cuatro factores que se consideran subyacentes en la enseñanza: la demanda que el área plantea a las prácticas del aula y el discurso relativo; el andamiaje para apoyar la enseñanza de los contenidos de lengua; las representaciones de los contenidos y el uso que se hace de ellos, y el ambiente del aula (MET Project, 2010f).

El sistema identifica 13 elementos que constituyen dimensiones independientes, y permite evaluarlos con una rúbrica, en una escala de 1 a 4:

- Propósito
- Desafío intelectual
- Representación del contenido
- Conexiones con el conocimiento previo
- Conexiones con la experiencia personal
- Modelamiento
- Estrategias explícitas de enseñanza
- Práctica guiada
- Discurso que se tiene en el aula
- Instrucción basada en textos
- Lenguaje académico
- Manejo de las conductas
- Manejo del tiempo

El sistema prevé observaciones independientes de 15 minutos cada una durante una clase: dos en clases de 45 minutos y tres en las de 90. La investigación

sobre la herramienta es más reducida que en los casos anteriores, y se ha desarrollado principalmente en escuelas de Nueva York (MET Project, 2010d y 2010f).

Sistema QST (Quality of Science Teaching)

En este caso se trata de un nuevo protocolo de observación que está en desarrollo, en el marco del Proyecto sobre Medición de la Efectividad de la Enseñanza (MET Project).

El punto de partida para el desarrollo del QST es un sistema (The Teaching Event) para evaluar aspirantes a ocupar plazas de maestro en el sistema educativo de California, que comprende 13 áreas de enseñanza, mediante un portafolios con evidencias de la práctica docente que incluyen planes de clase, un video, muestras del trabajo de los alumnos y una reflexión del aspirante mismo. El sistema estima cinco dimensiones —planeación, enseñanza, evaluación, reflexión y uso de lenguaje académico— a partir de las cuales se espera derivar constructos para preparar protocolos y guías de codificación para observar sesiones de clase en aula (MET Project, 2010d).

Observaciones con grabación en video

Un tipo especial de observación es el que se hace no directamente en el aula, sino en forma diferida, sobre registros de sesiones de clases que tengan la fidelidad suficiente para permitirlo, por ejemplo, con grabaciones de video.

La ventaja de este tipo de trabajos —en particular en gran escala, cuando involucra a decenas o centenas de escuelas en regiones o incluso en países diversos— es que es más factible contar en esos lugares con técnicos que operen una cámara de video de manera eficiente, que con observadores con la calificación necesaria para obtener información válida y confiable. Es el caso de los trabajos que usan grabaciones en video de clases conducidas por maestros, y un ejemplo destacado de ellos es el “estudio de clases grabadas en video” que formó parte del Tercer Estudio Internacional sobre Matemáticas y Ciencias (TIMSS, por sus siglas en inglés) en 1995, y de su repetición en 1999 (TIMSS-R) (Stigler, Gallimore y Hiebert, 2000).

Otras ventajas son que se puede detener la reproducción de la grabación o repetirla mientras se califica, y que se puede volver a calificar con otros jueces y utilizando otros protocolos o escalas (*cfr.* MET Project, 2010g).

En el caso del TIMSS y el TIMSS-R sólo se grabó una clase por profesor, lo que obviamente no basta para tener una muestra suficiente de su forma de trabajar a lo largo de un curso. Por otra parte, las grabaciones en video, como la observación directa en vivo, no pueden captar lo que están pensando el maestro o los alumnos, sin lo cual no puede entenderse cabalmente el sentido de sus acciones. Esos aspectos “subjetivos” deberán atenderse con otros acercamientos.

El interés de los dos estudios no era, desde luego, la evaluación de maestros en lo individual, sino la identificación de características que distinguieran las formas de enseñar en culturas diferentes, por lo que se grabó una sola clase, pero de un número de maestros en cada país participante que pudiera tener cierta representatividad. En el caso del TIMSS se grabaron clases de 100 maestros en

Alemania, 50 en Japón y 81 en los Estados Unidos. En el TIMSS-R los países participantes fueron siete (Australia, los Estados Unidos, Holanda, Hong Kong, Japón, la República Checa y Suiza), y se trató de grabar una clase de matemáticas o de ciencias en 100 escuelas de cada país. Con ello fue posible detectar rasgos culturales interesantes (Stigler, Gallimore y Hiebert, 2000).

Una desventaja de la grabación en video se deriva de que el ángulo que puede captar una cámara convencional es reducido, por lo que muchas cosas que pasan en el aula no quedan registradas fuera de las que ocurren en ese reducido campo de visión. La forma de subsanar esta limitación es usar simultáneamente varias cámaras. Además de aumentar los costos, esta solución hace también que la interferencia producida por la observación misma sea muy grande, lo que pone en riesgo la validez.

En el marco del proyecto MET se explora el uso de videograbaciones mediante un sistema digital avanzado que, a diferencia de las cámaras convencionales, permite registrar lo que ocurre en el aula con una visión panorámica de 360°, y luego, al reproducir la grabación, hacer los acercamientos que se quiera.

Sin minimizar las ventajas que ofrecen las poderosas tecnologías digitales a las que hoy se tiene acceso, debe reiterarse también que su uso no evita la necesidad de contar con observadores calificados, así como con esquemas de categorías que permitan estructurar lo observado según las dimensiones pertinentes.

Para valorar adecuadamente el potencial de los acercamientos a la práctica docente basados en distintos tipos de grabación de video sería necesario analizar en detalle tanto los pros y los contras como el costo y los beneficios de las distintas opciones, desde los videos con cámaras y micrófonos múltiples, y uno o incluso dos camarógrafos, hasta la grabación en 360°, cuya efectividad debe aún demostrarse, para ver si puede dar efectivamente información de mejor calidad que las otras opciones a menor costo o no.

Para terminar este apartado, señalaremos que se suele pensar que el mejor acercamiento a la práctica docente es siempre la observación, y que el único obstáculo para su uso es el costo que supone.

Sin minimizar el potencial de esta estrategia de obtención de información, se deben advertir al menos dos inconvenientes prácticos que tiene: el derivado de la dificultad de obtener información confiable sobre conductas complejas que no se prestan para el uso de protocolos estructurados y deben basarse en registros fundamentados en inferencias que hacen observadores desigualmente preparados, y que el mejor observador no puede saber lo que los actores piensan, por lo que hay que interrogarlos.

Acercamientos basados en análisis de productos

Los acercamientos al estudio de las prácticas docentes de los dos apartados anteriores se basan en lo que reportan los maestros mismos o en la observación de su trabajo. Otros acercamientos se basan en el análisis de los productos de las prácticas, como planes de clase, cuadernos de alumnos, exámenes aplicados o tareas encomendadas.

Se trata, pues, de una tercera categoría, distinta tanto de los trabajos basados en lo que los docentes mismos reportan (sea en forma espontánea, en bitácoras, diarios u otro tipo de autorreportes o en respuesta a las preguntas que se les dirigen por medio de cuestionarios), como de los que utilizan cualquier forma de observación.

Análisis de tareas

Un primer acercamiento de este grupo consiste en el análisis de las tareas que los alumnos llevan a cabo por indicación del maestro. Es claro que la revisión de esos trabajos podrá revelar muchas cosas sobre la práctica del maestro, como la forma en que entiende su papel, el dominio que tiene de los contenidos, su concepción de la evaluación, el grado en que ofrece retroalimentación más o menos rica a los alumnos, entre otras. Por ello se entiende que un ejemplo temprano de este tipo de acercamiento lleve como subtítulo "Abriendo una ventana sobre las prácticas del aula" (Matsumura y Pascal, 2003).

Los trabajos en que se basa este aparatado, desarrollados en el National Center for Research on Evaluation, Standards and Student Testing (CRESST), con sede en la Universidad de California en Los Ángeles, definieron la calidad de las tareas según tres dimensiones: el nivel de demanda cognitiva que representaban, la claridad de los objetivos de aprendizaje a los que una tarea debería contribuir, y la claridad con que se especificaban los criterios de calificación (Matsumura y Pascal, 2003).

Las puntuaciones asignadas en una escala de cuatro puntos en cada una de estas dimensiones se combinan para formar una puntuación de calidad global.

De cada maestro participante (181 de 35 escuelas, de los grados 4°, 7° y 10°) se recogieron materiales de tres tareas (dos sobre comprensión lectora y una sobre expresión escrita), incluyendo información sobre la planeación, las instrucciones dadas a los alumnos, las tareas entregadas por cuatro alumnos (dos de alto y dos de medio rendimiento), y se observó a los maestros en clase dos veces.

Se analizaron las tareas recolectadas y los resultados se sintetizaron para formar indicadores de la práctica docente. Se encontró un nivel aceptable de consistencia entre los analistas, siendo necesarias tareas de tres o cuatro alumnos de cada maestro para tener estimaciones estables, y sólo si las tareas habían sido diseñadas por el maestro y no tomadas de otra fuente.

La calidad de las tareas se asoció con la de la enseñanza, según las observaciones de la misma, y también con la calidad del trabajo de los alumnos; aquellos estudiantes cuyo maestro dejaba tareas cognitivamente más exigentes y manejaban criterios de calificación más claros mostraron también mayor avance en evaluaciones externas. Sin embargo, en general la calidad de las tareas asignadas por los maestros no fue muy alta (Matsumura y Pascal, 2003).

Los portafolios y sus variantes

Los portafolios son una herramienta ampliamente utilizada desde hace tiempo en sistemas de evaluación de alumnos, pero también de maestros, escuelas, etcétera; se inspiran en las carpetas en que pintores y otros artistas gráficos reúnen muestras de su obra para que los posibles clientes se formen una idea de su producción.

En el caso de la evaluación de maestros, los portafolios consisten también en conjuntos de evidencias de su trabajo, como planes de clase, cuadernos de alumnos, registros de clases impartidas, entre otros. Los portafolios contienen materiales seleccionados por el sujeto mismo que va a ser evaluado, del que se espera una reflexión sobre su propia situación, una *autoevaluación* a partir de los materiales incluidos en el portafolios, que luego serán valorados por otras personas, las cuales podrán ser autoridades, pares u otros evaluadores externos.

La importancia de los portafolios como herramientas para la evaluación de los docentes se aprecia con facilidad si se considera que gran parte de los sistemas que se han desarrollado con ese propósito la incluyen.

Es el caso de los sistemas de indicadores múltiples a los que se hizo referencia en un apartado anterior: el del National Board for Professional Teaching Standards, el del Interstate New Teacher Assessment and Support Consortium, y el sistema Praxis del Educational Testing Service.

Además de lo anterior, los sistemas educativos de varios estados y distritos escolares de los Estados Unidos han desarrollado sus propios sistemas de evaluación docente basados en portafolios, como los siguientes:

- Washington ProTeach Portfolio
- Performance Evaluation Program, del distrito de escuelas públicas de la ciudad de Alexandria, en el estado de Virginia
- Teaching Event, sistema del Performance Assessment for California Teachers
- Kansas Performance Teaching Portfolio (*cf.* Potemski et al., 2011: 19-21)

La “Guía de productos para evaluación de maestros” (“Guide to Teacher Evaluation Products”), que mantiene el National Comprehensive Center for Teacher Quality en su página web (www.tqsource.org), contiene información sobre otras herramientas de tipo portafolio, así como de otras de las consideradas en este documento como medidas del desempeño de los alumnos y modelos de valor agregado, encuestas de alumnos, autorreportes de maestros, protocolos de observación y artefactos de enseñanza.

La “Guía” contiene información bastante amplia de cada instrumento, incluyendo la descripción de cada uno así como referencias bibliográficas relevantes y datos de la institución o las personas que lo han desarrollado, con páginas web o correos electrónicos de contacto (NCCTQ, 2011).

La ya larga experiencia sobre el uso de portafolios ha mostrado su valor como herramienta de desarrollo profesional, cuando se enfatiza la dimensión formativa de la evaluación que se puede hacer mediante ésta.

Al usar un portafolio con el propósito de observar/medir/evaluar la práctica docente, como ocurre con los demás instrumentos revisados, se aplica la necesidad de precisar el foco de atención al que se refiere éste: en el caso que nos ocupa, el aspecto o los aspectos de la práctica que se quiere observar. Lo mismo en cuanto a la necesidad de explicitar las categorías que se usarán para sistematizar lo observado, así como la forma de pasar a la medición y al juicio evaluativo.

Hay ejemplos destacados de sistemas de evaluación de maestros basados en la preparación de portafolios y en el análisis de su contenido, a los que se ha hecho ya alusión en el apartado en que se sintetizan tres revisiones de literatura.

Otras herramientas que constituyen variantes de portafolios son las siguientes.

Sistema IQA (Instructional Quality Assessment)

Es un desarrollo del trabajo de Matsumura sobre el análisis de las tareas hechas por los alumnos como forma de aproximarse al trabajo del maestro. Este nuevo sistema está en proceso, y centra la atención en Comprensión Lectora y Matemáticas en el nivel medio del sistema educativo.

El sistema comprende protocolos para evaluar la enseñanza mediante observaciones en aula, así como la calidad de las tareas asignadas por el maestro. Se reportan trabajos para estudiar la confiabilidad y la potencial validez de la información obtenida.

Esos trabajos coinciden con otros en encontrar una amplia variación de la calidad de la enseñanza en las áreas estudiadas, con un nivel promedio no muy alto. La calidad de la información fue mejor en Matemáticas que en Lectura. Los resultados de la aplicación del IQA predicen los obtenidos por los alumnos en algunas evaluaciones externas, después de controlar otros factores posiblemente asociados. La solidez de los resultados es limitada porque se derivan de un número reducido de casos (Matsumura *et al.*, 2006).

Paquetes de materiales (Artifacts Packages)

Este tipo de instrumentos es una variante de portafolios y no se limita a la revisión de tareas, sino que incluye cualquier otro tipo de material (*artifact*) que informe sobre las prácticas que tienen lugar en el aula: planes de clase; notas que el maestro entrega a los alumnos; fotografías de la forma en que se organiza a los alumnos para trabajar o de lo que escribe en el pizarrón; trabajos de los alumnos, sean o no tareas; grabaciones de sesiones; etcétera.

Para entender el acercamiento es útil analizar la expresión que el instrumento usa para referirse a la forma de proceder: *scoop notebook*. El segundo término alude simplemente a que los materiales recogidos se ponen en una libreta o carpeta. Interesa más el otro término: *scoop*, que quiere decir "red", en particular una para atrapar mariposas u otros insectos, o para sacar los peces de una pecera. Así como un biólogo se dedica primero al trabajo de campo, durante el cual recopila el mayor número posible de especímenes que analizará después en el laboratorio, así los investigadores de la práctica docente pueden primero dedicar tiempo a recolectar todas las evidencias que puedan (materiales, *artifacts*) del trabajo de docentes y alumnos, para luego estudiarlas con detalle.

El propósito del proyecto *scoop* fue desarrollar un acercamiento alternativo al estudio de la práctica docente empleando artefactos y materiales para alcanzar a hacer una representación de dicha práctica suficiente a fin de que una persona, sin observar directamente al maestro ni al aula, pudiera hacer juicios válidos sobre algunos aspectos particulares únicamente con base en esos materiales.

Además de incluir instrucciones para los maestros sobre cómo reunir los materiales, un componente clave es el que constituyen las guías de calificación que

los jueces utilizan para analizarlos. El trabajo desarrollado se refiere en particular a las áreas de matemáticas y ciencias (Borko, Stecher y Kuffner, 2007).

Con base en estudios previos, se definieron diez dimensiones de las prácticas de enseñanza en las dos áreas en las que se pudiera apreciar su consistencia con las orientaciones para una enseñanza innovadora, según los estándares nacionales de Matemáticas y Ciencias. En el caso de Matemáticas las dimensiones son:

- Tipos de agrupamiento utilizados para el trabajo de los alumnos
- Estructura de la clase
- Uso de formas diversas de representar las ideas y los conceptos
- Uso de herramientas matemáticas
- Demanda cognitiva que se maneja
- Presencia de un discurso matemático común
- Manejo de explicaciones y justificaciones
- Manejo de solución de problemas
- Formas de evaluación utilizadas
- Uso de conexiones con la vida real o aplicación de conocimientos

Las dimensiones del área de ciencias son similares y, para cada una, se ofrece una rúbrica que define con precisión lo que se entenderá por una práctica de calidad alta, media o baja.

Las guías elaboradas a partir de esas dimensiones (una para Matemáticas y otra para Ciencias) se pueden usar tanto para hacer observaciones en aula, como para analizar los materiales recogidos en la “carpeta-red”, que los maestros reúnen mediante un trabajo intensivo durante una semana siguiendo las directrices que se les dan.

Es importante destacar que no se trata de una elección —arbitraria o no— hecha por cada maestro, sino de una recolección que busca representatividad en la materia de que se trate y en la semana indicada. Sí se pide al maestro que incluya también su punto de vista o sus consideraciones sobre los materiales recopilados, sobre el propósito de manejarlos y el uso que les daba. Los materiales incluyen unos que se manejan antes de la clase (planes, notas, rúbricas para evaluar), durante la clase (textos escritos en el pizarrón, materiales proyectados, trabajos de los alumnos) y después de la clase (tareas, exámenes, ejemplos de portafolios).

Aunque se trata de un trabajo en desarrollo, los resultados muestran que los juicios sobre la práctica docente basados en el sistema tienen niveles aceptables de confiabilidad y validez (Borko, Stecher y Kuffner, 2007; Borko *et al.* 2005).

Al no basarse ni en la información proporcionada por las personas mismas que son objeto de estudio, ni en la que pueden obtener observadores independientes, las herramientas que buscan estudiar la práctica docente a partir de la revisión de sus productos son una categoría diferente de los acercamientos a las prácticas basados en la observación o en la interrogación de los maestros mismos.

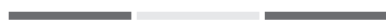
Como esos otros acercamientos, los que se basan en el análisis de productos de las prácticas tienen pros y contras, por lo que unos y otros deberían verse como enfoques complementarios y no alternativos.

La utilización de productos de las prácticas como fuentes de información comparte con la observación la ventaja de que no se ve afectada por el riesgo de captar lo deseable más que lo real; como la observación de grabaciones de observaciones, los productos de las prácticas pueden ser revisados una y otra vez por distintos analistas y utilizando protocolos diferentes.

Estos acercamientos también comparten con la observación su incapacidad de captar lo que pasaba por la mente de los maestros al realizar las actividades de las que se derivaron los productos, lo que sólo se podrá explorar a través de la versión que den los actores estudiados respondiendo a preguntas dirigidas a ellos. ●

Capítulo 3

La evaluación de los maestros y la calidad educativa



La distinción entre evaluación de enfoque formativo y la de sumativo es clásica, desde su introducción por Scriven (1967). En lo relativo a la evaluación del aprendizaje de los estudiantes, la distinción fue elaborada por Bloom (1968 y 1971), así como por Sadler (1989) y otros autores. Hoy se considera que la evaluación formativa es una parte fundamental de los enfoques pedagógicos de tipo constructivista, derivados de la revolución cognitiva (cfr. Martínez Rizo, 2012a).

De acuerdo con esos enfoques, el propósito más importante de la evaluación del aprendizaje no es sancionar el nivel alcanzado al final de un ciclo en términos de satisfactorio o no en relación con las metas que define el currículo (propósito sumativo), sino el de contribuir al logro de esos propósitos de aprendizaje al dar al docente y a los estudiantes mismos una retroalimentación que, por su oportunidad y riqueza, les permita orientar mejor sus esfuerzos ulteriores (propósito formativo).

La distinción es asimismo pertinente en el caso de la evaluación de los maestros, cuyo propósito puede ser sólo identificar a los que alcanzan ciertos estándares de calidad y a los que no logran hacerlo, pero también puede servir para ayudar a los que aún no lo hacen a orientar mejor sus esfuerzos para conseguirlo.

A la distinción entre evaluación *del* aprendizaje o evaluación *para* el aprendizaje (*assessment of learning* vs. *assessment for learning*) corresponde la que opone la evaluación *de* la enseñanza a la evaluación *para* la enseñanza (*assessment of teaching* vs. *assessment for teaching*) (Shulman, 2009: 234-244).

Es claro que la evaluación por sí misma no es interesante, si no va unida de alguna manera al mejoramiento. Para que lo último ocurra, parece necesario que, sin dejar de ser objetiva, la evaluación incluya un elemento “amigable” de apoyo y consejo por pares reconocidos y capacitados.

Esta concepción de la evaluación es congruente con una visión de la mejora de las escuelas que considera que dicha mejora sólo se podrá conseguir mediante estrategias basadas en la experiencia y el trabajo profesional de los maestros, quienes pueden hacer que las escuelas se vuelvan organizaciones efectivas, centradas en el aprendizaje, y capaces de mejora continua.

La mejora de las escuelas, en cambio, no puede lograrse solamente gracias a apoyos externos que se ofrezcan a un cuerpo docente poco profesional que sólo sea capaz de ejecutar en forma rutinaria tareas simples prescritas con detalle.

Una consecuencia más de esta manera de concebir la evaluación es la que utiliza la idea de *oportunidad de enseñar* (*opportunity to teach*), similar a la oportunidad de aprendizaje (*opportunity to learn*): para que el trabajo del maestro sea efectivo debe haber ciertas condiciones en las escuelas, incluyendo presencia de recursos suficientes y ausencia de presiones o condicionamientos adversos, lo que implica apoyo de las instancias superiores, desde el nivel de zona hasta el más alto (Lanson-Billings, 2009: 206-222).

La evaluación por sí misma no produce mejora alguna; puede aportar información sobre la situación prevaleciente y, en el mejor de los casos, sobre los obstáculos que se oponen a la calidad y los factores que pueden favorecerla. Esto es valioso, pero la calidad misma implica mucho más. En lo que se refiere a los maestros, evaluar su desempeño no los hará mejores en forma automática; para que haya buenos maestros deberán estar presentes otros elementos:

- Una buena cantera de futuros docentes que se preparen en programas que den excelente formación inicial. Incluye la presencia de condiciones que hagan atractiva la profesión docente, cuidadosos procedimientos de selección e instituciones formadoras sólidas.
- Buenas evaluaciones para definir el ingreso a la profesión, con procesos de transición de la escuela al primer trabajo e inducción a éste, incluyendo el acompañamiento de tutores experimentados.
- Buenos sistemas de promoción horizontal y vertical de los maestros en servicio, con base en evaluaciones confiables y válidas de su desempeño, incluyendo la certificación de los más destacados.

De lo anterior se desprende que, para que pueda contribuir a mejorar la calidad educativa, un sistema de evaluación de docentes no se puede reducir a valorar el desempeño de los que están frente a grupo.

Un sistema completo de evaluación de docentes deberá comenzar desde mucho antes, para promover que accedan a la profesión jóvenes con cualidades que permitan esperar que serán buenos maestros, que reciban una buena formación para ello, y que sean apoyados al comenzar a trabajar en las escuelas y a lo largo de su trayectoria en el sistema educativo. Por ello un sistema integral de evaluación de maestros deberá incluir varios elementos.

PIEZAS DE UN SISTEMA INTEGRAL DE EVALUACIÓN DE MAESTROS

En un texto reciente, una de las expertas más reconocidas en el campo de la investigación sobre docentes, Linda Darling-Hammond, considera que un sistema completo de evaluación de maestros debería contemplar tres componentes, para atender en particular tres etapas de la vida profesional de un docente:

Tabla 4 Un continuo de evaluaciones del desempeño docente

Primer nivel	Segundo nivel	Tercer nivel
Para licencia inicial: <ul style="list-style-type: none"> ▪ Al entrar a la profesión ▪ Con altos estándares comunes a toda forma de acceso a la docencia ▪ Con evaluaciones del conocimiento de los contenidos ▪ Cuyo resultado se use para la acreditación de programas de formación 	Para licencia profesional: <ul style="list-style-type: none"> ▪ Después de inducción y antes de plaza definitiva ▪ Con evidencia sistemática de práctica docente y aprendizaje de los estudiantes ▪ Cuyo resultado se use también para informar el apoyo a los docentes y el desarrollo profesional 	Para certificación avanzada: <ul style="list-style-type: none"> ▪ Después de definitividad ▪ Evaluación de logros de docente experimentado ▪ Certificación del NBPTS o alternativa local ▪ Evidencia utilizada para pago diferencial y para acceso a puestos de dirección

Fuente: Darling-Hammond (2012: 11, fig. 1).

En un texto anterior, la misma autora (Darling-Hammond, 2008: 752) se refiere a un buen sistema de evaluación de docentes como un “banco de tres patas” (*three legged stool*), porque debe incluir tres componentes igualmente importantes que son indispensables para que el “banco” de la evaluación se mantenga en pie:

Acreditación de programas de formación inicial

Un sistema educativo sólo podrá contar con buenos docentes si hay programas de formación muy sólidos. En México este punto es crucial, dada la debilidad de la mayoría de las escuelas normales.

Partiendo del supuesto de que una buena preparación para la enseñanza, desde los niveles elementales, implica necesariamente una formación de nivel terciario, es posible pensar en más de una forma de acceder a la profesión docente: una es la usual en México, de una licenciatura enfocada expresamente a la docencia, pero puede haber otras, como una licenciatura en un área de conocimiento a la que siga una formación de posgrado para la docencia, y en todos los casos será fundamental una etapa de práctica supervisada por un docente con experiencia.

No hay razones académicas para descartar en principio alguna de estas formas, e independientemente de la que se adopte para organizar la preparación del futuro maestro, deberá haber mecanismos de evaluación que aseguren que todo programa formativo sea sólido, incluyendo la revisión de los procedimientos de reclutamiento y selección de los aspirantes, el contenido del programa formativo, la calidad de los formadores, los recursos de biblioteca y otros, las prácticas profesionales, etcétera.

Y como apunta la tabla de Darling-Hammond, los resultados de las evaluaciones del desempeño en la docencia de los egresados de un programa formativo deberán tener un peso significativo en la acreditación del programa mismo.

Licencias para el ingreso a la profesión

Una buena formación inicial es una condición necesaria, pero no suficiente, para tener un buen docente. Es frecuente escuchar a maestros decir que aprendieron más en la práctica que en la escuela normal.

Por ello el inicio de la vida profesional es una etapa de gran importancia para que el nuevo docente ponga en práctica lo que aprendió, en las circunstancias reales de una escuela y con alumnos concretos. La transición no es fácil, y si el novato llega a un medio en el que prevalezcan prácticas poco adecuadas y no cuenta con apoyo experto, es probable que se adapte a lo prevaleciente y olvide la teoría.

Una etapa que dure al menos dos o tres cursos escolares parece indispensable como periodo de aprendizaje y de prueba antes de tener una plaza de base. Las evaluaciones de esa etapa, como condición para recibir la licencia profesional o el nombramiento definitivo, deberán comprender varios elementos.

Es razonable que incluyan una valoración del conocimiento que tenga el maestro novato de los contenidos a enseñar, así como de los principios pedagógicos que deben sustentar su enseñanza, siempre y cuando esa valoración no se haga con base en instrumentos de escasa calidad, como cuando se reducen a la aplicación de una prueba compuesta por preguntas de respuesta estructurada.

Es necesario que se evalúen competencias cognitivas complejas y, en particular, el *conocimiento pedagógico de los contenidos*, del que ya se ha tratado en un punto anterior. Sin duda es importante considerar los niveles de aprendizaje que alcanzan los estudiantes atendidos por el maestro, pero además no puede estar ausente la valoración de las prácticas docentes mismas, lo que implica, como ya se ha apuntado, la utilización de instrumentos como los portafolios o la observación en aula.

Otro componente de una buena evaluación del desempeño de los nuevos docentes es la opinión calificada de directores y supervisores, con los cuidados debidos para evitar sesgos, y muy especialmente la valoración que haga el monitor experimentado que siempre debería acompañar a estos docentes.

Certificación de niveles avanzados de desempeño

Una vez que un docente ha accedido a la titularidad de una plaza —lo que no debería ocurrir antes de dos o tres años de ejercicio, y con base en evaluaciones como las mencionadas en el punto anterior—, su carrera puede desarrollarse en formas diversas. La investigación ha mostrado que la experiencia parece incidir significativamente en la forma de trabajar de los maestros durante los primeros años de vida profesional, pero que, en muchos casos, después de cuatro o cinco años no se aprecian cambios en las rutinas establecidas, que se mantienen sin transformaciones importantes durante el resto de la trayectoria.

Lo anterior parece congruente con la visión del maestro como técnico que ejecuta en forma rutinaria tareas preestablecidas, pero no resulta adecuada en una perspectiva del maestro como profesional, que debería actualizarse en forma permanente para enfrentar situaciones inéditas e incorporar los avances del campo de conocimiento de su ámbito profesional.

En la perspectiva tradicional tenían sentido sistemas escalafonarios basados en la antigüedad y la escolaridad; en la nueva perspectiva son necesarios sistemas de evaluación que consideren expresamente la práctica valorando la capacidad de los docentes para adaptarla a circunstancias y estudiantes cambiantes, e incorporando los avances de su campo como verdaderos profesionales.

Buenas evaluaciones de ese tipo —que, una vez más, tendrán que considerar el aprendizaje de los alumnos pero también la práctica misma, y nunca reducirse a pruebas simples— podrían ser la base de sistemas de estímulos económicos o pagos diferenciados que reconozcan de manera confiable a los mejores maestros, así como sustentar las decisiones relativas a la promoción de los maestros a niveles superiores de responsabilidad.

Pero además, en una perspectiva formativa, buenas evaluaciones darán valiosa información para la planeación de actividades de desarrollo profesional con las que los docentes se actualicen y se enriquezca su práctica.

RETOS PARA EL DESARROLLO DE BUENOS SISTEMAS DE EVALUACIÓN

Conviene distinguir tres tipos de problema que se deben enfrentar para desarrollar un buen sistema de evaluación de maestros: desafíos de caracteres conceptual, técnico y práctico (cfr. Wilson, 2009).

Desafíos conceptuales: ¿qué es un buen maestro?

El primer tipo de problemas es de naturaleza conceptual. El punto de partida de toda buena evaluación, como se ha dicho, es una definición precisa del estándar de referencia contra el que se contrastará la realidad a evaluar, para valorarla como adecuada o inadecuada, deficiente, satisfactoria o sobresaliente. En el caso de la evaluación de los maestros, hay consenso en el sentido de que el punto de partida debe ser una definición clara de lo que se entenderá por buen maestro.

En un texto ya citado, Linda Darling-Hammond recuerda el dicho del gato de Cheshire en *Alicia en el país de las maravillas*: “Si no sabes a dónde quieres ir, cualquier camino te llevará allá”, y a continuación dice:

Sin una clara comprensión de lo que los estudiantes deben aprender y de cómo puede apoyarlos para ello la enseñanza, es fácil vagar sin rumbo. Desarrollar estándares consistentes para la enseñanza, que comprendan toda la carrera del maestro e incorporen una visión compartida de las metas educativas y la instrucción que las apoyará es el primer paso (2012: 10).

En el punto de partida de la construcción de sistemas de evaluación deficientes, muchas veces se encuentran ideas confusas en cuanto a qué es lo que hay que medir. En nuestro caso, cuáles son los rasgos que deberá reunir un maestro para que se le considere bueno.

En una primera aproximación se puede pensar en dos tipos de definición:

- A partir de los resultados: un buen maestro es aquel cuyos alumnos muestran altos niveles de aprendizaje.
- A partir de los procesos: un buen maestro es aquel cuyo trabajo en el aula consiste en buenas prácticas docentes.

Una visión ingenua dirá que estas dos definiciones deberán llevar a las mismas conclusiones, identificando a las mismas personas como docentes buenos o no: si se escoge a los maestros que tengan buenas prácticas, seguramente sus alumnos aprenderán mucho, y si se escoge a los maestros cuyos alumnos aprenden mucho, seguramente las prácticas de los docentes serán buenas.

En realidad las cosas son mucho menos sencillas, por varias razones.

- En cuanto a los resultados del aprendizaje que alcancen los estudiantes, no es fácil llegar a un consenso sobre qué es lo deseable al respecto. Incluso si hay un currículo oficial que defina lo deseable, las áreas curriculares son varias, incluyen aspectos cognitivos y no cognitivos, los contenidos son potencialmente innumerables y, respecto a cada área, aspecto y tema, hay varios niveles de competencia a los que puede ser deseable que llegaran los alumnos de cierta edad o grado. Difícilmente un alumno tendrá buenos niveles de aprendizaje en todos los casos, y no es fácil decidir qué aspectos o combinación de aspectos se manejará como referente.
- Por lo que se refiere a los procesos es todavía más difícil precisar cuáles son las mejores prácticas que permitirán identificar a los buenos maestros; no bastan definiciones de tipo general que establezcan, por ejemplo, que un buen maestro hace trabajar en grupo a los alumnos, les hace preguntas de manera frecuente, o utiliza materiales didácticos variados. Las condiciones de los alumnos hacen que en unos casos sean mejores unas prácticas que no son adecuadas en otros. Si se piensa en el desarrollo de competencias complejas, está lejos de haber consenso en cuanto a cuáles son los mejores enfoques pedagógicos en la enseñanza de la lengua materna, una segunda lengua, las Matemáticas, las Ciencias Naturales o las Ciencias Sociales, por no hablar del desarrollo de actitudes y valores ciudadanos, etcétera.

La variedad de definiciones de buen maestro de las que parten algunos sistemas de evaluación es ilustrativa de la complejidad de la cuestión.

Según el National Comprehensive Center for Teacher Quality (NCCTQ), el Departamento de Educación de los Estados Unidos ofrece estas definiciones:

- **Maestro efectivo:** aquel cuyos estudiantes alcanzan tasas aceptables de crecimiento del aprendizaje (al menos un avance equivalente a un grado escolar, en un año académico).
- **Maestro altamente efectivo:** aquel cuyos estudiantes alcanzan tasas altas de crecimiento del aprendizaje (un avance entre un grado escolar y un grado y medio en un año académico) (Potemski *et al.*, 2011: 3).

El NCCTQ reconoce la importancia para la evaluación de los docentes de la información relativa al crecimiento del aprendizaje (*student growth data*), pero al mismo tiempo destaca la relevancia de "importantes aspectos adicionales de la enseñanza", y ofrece una definición que comprende cinco puntos:

- Un maestro efectivo tiene altas expectativas respecto a todos sus alumnos y les ayuda a alcanzar altos niveles de aprendizaje, según medidas de valor agregado, otras de crecimiento basadas en pruebas o formas alternativas.
- Un maestro efectivo contribuye a alcanzar resultados positivos académicos, actitudinales y sociales con sus alumnos, como asistencia regular, promoción oportuna al grado siguiente, graduación en el tiempo estipulado, autoeficacia y conducta cooperativa.
- Un maestro efectivo usa una gama de recursos para planificar y estructurar oportunidades de aprendizaje atractivas, monitorea con enfoque formativo el avance de los alumnos adaptando la enseñanza según sea necesario, y evalúa el aprendizaje utilizando múltiples fuentes de evidencia.
- Un maestro efectivo contribuye al desarrollo de aulas y escuelas que valoren la diversidad y un enfoque de ciudadanía.
- Un maestro efectivo colabora con otros docentes, administradores, padres de familia y otros profesionales de la educación para asegurar el éxito de los estudiantes, en especial de los que presentan necesidades educativas especiales o tienen más alto riesgo de fracaso (Goe *et al.*, 2008; Potemski *et al.*, 2011: 3).

Estos puntos son similares a los cinco de los estándares del National Board for Professional Teaching Standards (NBPTS) que ya se han presentado, y dicen que los buenos maestros deben:

- Estar comprometidos con los alumnos y su aprendizaje.
- Conocer las materias que enseñan y cómo enseñarlas a sus alumnos.
- Ser responsables de conducir y monitorear el aprendizaje de sus alumnos.
- Reflexionar sobre su propia práctica y aprender de la experiencia.
- Ser miembros de comunidades de aprendizaje (NBPTS, 1989).

Antecedente de esas y otras definiciones fue la lista de aspectos que Shulman y Sykes, en 1986, incluyeron en su revisión de la base de conocimientos para la enseñanza, en el trabajo que el grupo del proyecto sobre evaluación de maestros de la Universidad de Stanford elaboró para el Grupo de Trabajo sobre la Enseñanza como Profesión (Task Force on Teaching as a Profession) del Forum on Education and Economy de la Fundación Carnegie.

Con algunas modificaciones posteriores, las características que deberían reunir los maestros a juicio de esos autores incluyen:

- Educación general, competencias básicas en Lectura y Matemáticas.
- Conocimiento del contenido de las materias a enseñar.
- Conocimiento general de los principios pedagógicos.
- Conocimiento pedagógico de los contenidos.
- Conocimiento de concepciones educativas fundamentales.
- Conocimiento de los contextos en que se desarrolla la educación.
- Comprensión de la diversidad y las diferencias individuales de los alumnos.

- Conocimiento de los materiales, temas y alternativas del currículo.
- Habilidades prácticas (voz, estilo, equilibrio) (según Haertel, 1991: 10-11).

Estas enumeraciones de rasgos deseables de los buenos docentes no incluyen sus años de escolaridad o de servicio. Obviamente, de la formación se espera que los egresados tengan conocimientos suficientes en los aspectos mencionados, así como buena preparación para que desarrollen buenas prácticas, pero los listados enfatizan que lo que importa es que el maestro cuente con esos rasgos, sin importar cómo los adquirió. Si algunos docentes no muestran tener los conocimientos o dominar las prácticas de referencia, a pesar de haber pasado por un programa de formación inicial que se espera que los desarrolle, esa escolaridad no se puede considerar un rasgo positivo, y no deberá contar en una evaluación.

La complejidad de la conceptualización de un buen docente hace necesarias dos consideraciones más: que la definición que se utilice para desarrollar un sistema de evaluación debe partir de una amplia discusión de los actores relevantes, para llegar al consenso más amplio posible, y que las circunstancias particulares de algunos contextos de un sistema educativo harán necesario que se incluyan o destaquen ciertos elementos que en otros casos pueden no ser necesarios.

Esta postura no quiere decir que la definición del buen docente deba resultar de un proceso pseudodemocrático en el que todas las opiniones tengan el mismo peso independientemente de su consistencia, pero sí que un asunto tan delicado no puede basarse en lo que diga una persona o un grupo, aun si está formado por buenos especialistas. Es indispensable combinar los saberes de investigadores con los de maestros experimentados.

Por lo que se refiere a la necesidad de tener en cuenta la gama de situaciones en que tiene lugar la docencia, no se trata de esperar menos de los maestros que trabajen en contextos difíciles o atiendan a alumnos vulnerables. Se entiende que todos los niños tienen derecho a una buena educación, y que esta noción incluye el que las escuelas a las que asisten los más vulnerables deben tener condiciones iguales o mejores que las demás, incluyendo muy buenos maestros.

Lo que se subraya ahora es que no hay rasgos que definan al buen maestro *en general*, sino que hay muchas variantes de buen maestro: no es lo mismo, obviamente, un buen maestro de preescolar, uno de primaria o uno de secundaria y, en este último caso, uno de Lengua o uno de Matemáticas. Tampoco debe tener idénticas cualidades un maestro que trabaje en situación de multigrado que uno con alumnos de un solo grado, u otro en una escuela indígena bilingüe, etcétera.

Por eso el NBPTS no se limita a una definición general única, sino que ha desarrollado definiciones particulares para más de 30 tipos distintos de maestro, según áreas curriculares y niveles educativos y de edad. Otro tanto ha hecho el Sistema de Evaluación del Desempeño Profesional Docente (SEDPD) de Chile, seguramente el más completo de América Latina.

En relación con la necesidad de considerar el contexto para decidir qué elementos se incluyen en la definición del buen maestro, conviene reflexionar sobre una dimensión particular de la práctica docente, que coincide con lo que Pablo Latapí Sarre expresaba con la noción de "normalidad mínima", la cual incluye que

maestros y alumnos asistan con regularidad a la escuela, que se cumpla con los programas establecidos, que haya un ambiente de orden, etcétera.

En los sistemas educativos más consolidados, por ejemplo, no tendría sentido incluir entre los rasgos de un buen docente que nunca deje de asistir a la escuela con toda puntualidad, salvo en casos de fuerza mayor. En esos sistemas tal situación es simplemente impensable. En contextos como el mexicano, en cambio, es razonable sostener que el primer rasgo de un buen maestro es precisamente que nunca falte. Si un maestro se ausenta con frecuencia, todos los demás rasgos, por positivos que puedan parecer, se vuelven inoperantes.

Después de las consideraciones anteriores, y siguiendo el clásico esquema que distingue insumos, procesos y resultados (Modelo CIPP), en la tabla 5 se propone una sistematización de los componentes del perfil de un buen maestro. Después se presentarán posibles formas de obtener información respecto a cada elemento.

Tabla 5 Dimensiones de la definición de buen maestro

Insumos	Procesos	Resultados
Rasgos personales <ul style="list-style-type: none"> ▪ Capacidad intelectual ▪ Madurez afectiva 	Normalidad mínima <ul style="list-style-type: none"> ▪ Asistencia, puntualidad ▪ No dañar a los alumnos ▪ Cumplir el programa 	Formales <ul style="list-style-type: none"> ▪ Asistencia de alumnos ▪ Permanencia hasta el final ▪ Promoción de grado
Creencias, concepciones <ul style="list-style-type: none"> ▪ Sobre los alumnos ▪ Sobre él mismo ▪ Sobre los compañeros, la enseñanza, evaluación... 	Prácticas de enseñanza <ul style="list-style-type: none"> ▪ Planeación de clases ▪ Preparación de material ▪ Desarrollo de sesiones ▪ Interacción con alumnos 	Cognitivos <ul style="list-style-type: none"> ▪ Resultados de alumnos en todas las áreas ▪ Avance respecto a inicio ▪ Resultados de alumnos más vulnerables, equidad...
Conocimientos <ul style="list-style-type: none"> ▪ De competencias básicas ▪ Del currículo y contenidos ▪ Pedagógicos generales ▪ Pedagógico del contenido 	Prácticas de evaluación <ul style="list-style-type: none"> ▪ Monitoreo permanente ▪ Evaluación final ▪ Retroalimentación 	No cognitivos <ul style="list-style-type: none"> ▪ Salud física y mental ▪ Competencia ciudadana ▪ Responsabilidad y en general valores éticos
Actitudes y valores <ul style="list-style-type: none"> ▪ Reflexión, autocrítica ▪ Trabajo en equipo ▪ Comprensión de contexto ▪ Respeto de diversidad ▪ Valores éticos en general 	Prácticas de gestión <ul style="list-style-type: none"> ▪ Manejo de aula y grupo ▪ Trabajo en colectivo ▪ Relación con padres 	

Fuente: elaboración propia.

Desafíos técnicos: ¿cómo medirlos?

El segundo tipo de retos que debe enfrentar todo sistema de evaluación de los docentes se refiere a los de naturaleza técnica, que tienen que ver con la medición del grado en que el trabajo de un maestro presenta ciertos rasgos.

Debe recordarse que evaluar implica medir pero no se reduce a ello. Además de la medición del rasgo en cuestión, evaluar implica precisar el estándar de referencia que define el deber ser y comparar lo medido con ese estándar, para llegar al juicio de valor en el que consiste la evaluación.

Una vez precisados los elementos de la definición de lo que debe ser un buen maestro en listados de estándares como los presentados en el punto anterior, hay que pasar a la etapa de *medición* de esos rasgos en los evaluandos.

En seguida se presentan ideas en cuanto al momento y la forma de evaluar los componentes de la definición de buen docente sistematizados en la tabla anterior; en la siguiente se retoman esos componentes en los renglones, y en las columnas se distinguen tres etapas de la evaluación, siguiendo a Linda Darling-Hammond (tabla 4), durante la formación, al inicio del ejercicio y en etapas avanzadas de éste:

Tabla 6 Acercamientos a la evaluación de características del buen docente

	Durante la formación inicial	Al inicio del ejercicio profesional, antes de plaza de base	Etapas avanzadas del ejercicio, para estímulos y promociones
INSUMOS			
Rasgos personales	x	x	—
Creencias-concepciones	x	x	—
Conocimientos	x	Pruebas complejas de aspectos de este rubro	Pruebas de conocimiento pedagógico y contenidos
Actitudes y valores	x	x	—
PROCESOS			
Normalidad mínima		Director, control administrativo	Director, control administrativo
Prácticas de enseñanza		Portafolios evidencias, encuestas de alumnos, observación en aula de monitor, juicio del director	Portafolio, bitácoras, encuestas de alumnos, observación en aula de supervisor, director
Prácticas de evaluación			
Prácticas de gestión			
RESULTADOS			
Formales		Control administrativo	Control administrativo
Cognitivos		Resultados de alumnos en buenas evaluaciones (indicativo)	Resultados de alumnos en buenas evaluaciones (indicativo)
No cognitivos		Resultados según apreciación de director y padres (indicativo)	Resultados según apreciación de director y padres (indicativo)

Fuente: elaboración propia.

En páginas anteriores se ha visto la amplia gama de nuevos acercamientos que se han desarrollado en las últimas décadas, pero también se habrá podido valorar la dificultad que hay para obtener información de buena calidad con las nuevas herramientas, dada la complejidad de la práctica docente, que es multidimensional y dinámica, ante la presencia de numerosos factores que inciden en los resultados de los alumnos además de la enseñanza, y dado que en muchos casos la medición se basa en el juicio subjetivo de supervisores, directores o calificadoros.

Algunos de los trabajos recientes más notables por desarrollar mejores sistemas de evaluación de los docentes, como el del National Comprehensive Center

for Teacher Quality, o el Proyecto MET (Measures of Effective Teaching), coinciden en dar prioridad a los resultados del aprendizaje de los alumnos para la evaluación de los docentes, con la justificación ya mencionada, de que si el propósito último de la enseñanza es el aprendizaje, bastaría en principio medir lo que alcanzan los estudiantes de un maestro para valorar la calidad de la enseñanza, sin tener que observar la práctica docente. Ya se han señalado los problemas de esta idea, que se pueden reducir a dos:

- La dificultad de tener buenas mediciones del aprendizaje de gran número de alumnos, que no se reduzcan a áreas y temas que se prestan para ser medidos con pruebas estandarizadas, sino que incluyan todas las áreas del currículo, así como todos los niveles de complejidad cognitiva.
- La dificultad de atribuir a la enseñanza de un maestro en particular los resultados del aprendizaje de ciertos alumnos, controlando bien los demás factores que inciden en esos resultados.

La dificultad práctica del primer punto es muy grande, pero en el segundo caso el problema no es sólo práctico, sino de principio. Se trata de establecer con solidez una relación causal entre el aprendizaje de ciertos alumnos y la enseñanza de un maestro. Definir al buen maestro como *aquel cuyos alumnos tienen buenos niveles de aprendizaje*, según las propuestas del Departamento de Educación citadas antes, es claramente insuficiente.

Lo adecuado sería decir que un buen maestro es *aquel cuyos alumnos tienen buenos niveles de aprendizaje gracias a su enseñanza*, y para comprobar esto no bastan correlaciones simples ni, en sentido estricto, resultados de modelos de valor agregado.

En el texto más completo que hay actualmente sobre validez, según una opinión tan calificada como la de Brennan (2013), Kane (2013) se refiere a la dificultad que hay para validar interpretaciones de los resultados de los test en términos causales.

Recordando el riesgo de caer en razonamiento circular al interpretar en términos causales una correlación, Kane presenta los requisitos que se deben cumplir para validar tales interpretaciones; añade que se trata de una clase muy fuerte que sin duda es deseable hacer, pero que es muy difícil justificar (43), y precisa:

Así pues, parece que, en la mayoría de contextos, estamos muy lejos de contar con interpretaciones de los resultados de test que se basen en modelos causales robustos, y que la validación de tales interpretaciones causales está todavía más lejos (44).

Los trabajos que dan prioridad a los resultados de aprendizaje para evaluar la efectividad de los docentes utilizan esos resultados (puntajes en pruebas con modelos de valor agregado) como criterio para validar otros instrumentos basados en observaciones de la práctica docente, en encuestas entre alumnos o maestros, o en la revisión de materiales con evidencias de la práctica en aula.

Cuando no se encuentra una clara relación entre los resultados del aprendizaje y los datos que aportan esos otros instrumentos, la conclusión es que estos últimos no están midiendo bien la efectividad de los docentes. No se plantea la cuestión de la calidad de los datos que se obtienen sobre el aprendizaje de los alumnos, de la que hay razones para dudar, como ya se ha mencionado.

Lee Shulman dice al respecto:

Con respecto a los abusos de los modelos de valor agregado y formas similares de evaluación de la calidad, cuya utilización puede carecer de soporte sólido, e incluso ser desorientadora, ya no podemos seguir fingiendo que no nos damos cuenta (2009: 243).

Lo anterior no implica que siempre sea clara la validez de los resultados de otros acercamientos para valorar la efectividad de los maestros; toda herramienta tiene puntos fuertes y débiles, pero algunas opiniones negativas se deben matizar.

En este sentido es útil citar de nuevo a Shulman, quien dice lo siguiente en relación con los portafolios:

Algunos especialistas en medición nos habían advertido que los portafolios nunca funcionarían bien, porque "los maestros harán trampa". ¿Qué quiere decir para ustedes hacer trampa?, pregunté yo, y me respondieron: los maestros van a hablar entre sí y a ayudarse para preparar su portafolios. ¡Aleluya!, dije yo. De lo que ustedes se quejan es de que los maestros probablemente aprenderán unos de otros al prepararse para las evaluaciones (241).

El párrafo anterior cobra todo su sentido si se tiene presente la distinción entre evaluación de enfoque sumativo y formativo, entre evaluación de la enseñanza y evaluación para la enseñanza. Por otra parte, la postura de Shulman no es ingenua; reconoce que: "algunas personas simplemente no entienden la frontera entre una asesoría legítima y hacerle el trabajo a otras personas. Pero ése es un problema que todo director de tesis debe enfrentar en innumerables ocasiones" (241).

Shulman recuerda que cuando comenzó el trabajo para desarrollar el sistema del NBPTS, y buscando la asesoría más calificada para cuidar la calidad técnica de los instrumentos a desarrollar, invitaron "al Zeus de la psicometría", Lee Cronbach, quien aceptó dejar su retiro y trabajar con el equipo, con una advertencia:

Trabajaré en este proyecto con la condición de que ustedes se pregunten primero qué tipo de evaluaciones serán más fieles a la forma en que entienden la enseñanza y el aprendizaje, y es más probable que sean de valor para el campo. Entonces mi tarea será pensar cómo hacer que esas evaluaciones sean viables en términos psicométricos. El día en que yo vea que ustedes corrompen lo que estén haciendo para atender algún principio o práctica psicométrica, en ese momento regresaré a mi retiro. No podemos permitir que el rabo metodológico sea el que dirija al perro de la buena práctica pedagógica (Shulman, 2009: 240).

Desafíos prácticos: ¿cómo implementar las medidas?

Este tipo de retos tienen que ver con las dificultades especiales que se presentan cuando se trata de evaluar un gran número de sujetos, además de los problemas que se derivan de la complejidad de los aspectos a evaluar y de la posibilidad de hacerlo mediante indicadores que se presten más o menos a ser observados y registrados de manera confiable y válida.

Además, puede haber problemas importantes derivados de la oposición de los sujetos a evaluar. En ocasiones el rechazo tiene sustento porque se refiere a evaluaciones mal diseñadas; en otras se trata de posturas radicales que se oponen a cualquier forma de evaluación, pero en todas se trata de obstáculos que se debe tener en cuenta.

Un tercer tipo de desafío práctico, que es fundamental tener en cuenta, es el que se refiere al tiempo que implica desarrollar un buen sistema de evaluación de maestros. Ya se ha citado la opinión del primer director del Consejo del NBPTS, James Kelly, en el sentido de que “establecer plenamente un esquema nacional de certificación llevaría 20 años” (Ingvarson y Hattie, 2008: 3). El sistema chileno, que pudo aprovechar la experiencia estadounidense, llevó diez años para consolidarse.

No es frecuente que las autoridades superiores de los sistemas educativos sean conscientes de la importancia de respetar plazos de tales dimensiones, dado que los tiempos políticos suelen ser mucho más cortos.

Debe tenerse en cuenta, además, que si el tiempo necesario para desarrollar un buen sistema de evaluación docente es largo, todavía lo es más el necesario para elevar en un grado significativo la calidad promedio de los maestros de cualquier sistema educativo grande. Conviene destacar lo anterior porque a la visión de corto plazo de algunos actores políticos se añade la expectativa ingenua de ver resultados notables de mejora muy poco tiempo después de llevar a cabo una evaluación, sin tener en cuenta que evaluar por sí mismo no mejora las cosas, y que para ello se necesita mucho más.

Ni siquiera en el mediano plazo es realista esperar mejoras sustanciales en las prácticas de maestros con bastantes años de ejercicio y que arrastran deficiencias desde antes de entrar a la escuela normal, a las que se han añadido vicios añejos presentes en muchas escuelas. Si se trata de maestros más jóvenes la tarea es más factible, pero sólo a condición de poner en marcha programas sólidos de desarrollo profesional y de ofrecer a las escuelas apoyo material y pedagógico sustancial, todo lo cual es caro e implica años de trabajo.

Mejorar significativamente la calidad de los nuevos maestros que egresen de los programas de formación inicial, por su parte, además de que también es complejo y costoso, no dará resultados en el corto plazo, ya que los nuevos docentes irán sustituyendo a los anteriores poco a poco, y no se puede descartar que los nuevos adopten costumbres añejas viciadas, en vez de sustituirlas por prácticas mejores.

Por ello no basta establecer estándares ambiciosos para la práctica docente; en más de una ocasión se ha visto que después de hacerlo, en algunos sistemas educativos, los sistemas de evaluación se han relajado para no enfrentar la presencia

de un buen número de docentes que no satisfacen los estándares (cfr. Darling-Hammond, 2008: 772-773).

En este sentido son pertinentes dos consideraciones más, que se desarrollan en los dos incisos siguientes.

a) El problema de los niveles altos de demanda cognitiva

No es raro que personas no familiarizadas con la docencia en los niveles básicos de la educación consideren que se trata de una tarea sencilla que puede llevar a cabo correctamente todo aquel que haya terminado la educación media superior o cualquier licenciatura. Esta idea sustenta programas que buscan llevar a las aulas de primaria, en particular en un medio desfavorable, a egresados universitarios dispuestos a dedicar un tiempo a una tarea emprendida muchas veces con generosidad y la más noble intención.

La misma idea implica descalificaciones injustas expresadas en términos de “el que puede, hace, el que no puede, enseña”.

En uno y otro caso el supuesto implícito es que la docencia en educación básica no sería una actividad profesional, sino una para la que bastaría una formación general y buena voluntad, sin preparación específica.

Es posible que un egresado universitario muy capaz y dedicado consiga mejores resultados, en algunos casos, que algunos egresados de instituciones formadoras de maestros con formación deficiente, y con dudosa motivación. Pero lo que hoy se sabe sobre la dificultad que representa desarrollar competencias complejas con niños de contexto vulnerable muestra que una buena docencia en los niveles básicos de la educación es una tarea profesional que requiere de una formación de nivel universitario, con periodos de práctica profesional supervisada como los que hay en otras profesiones, que permita dominar un cuerpo de conocimientos y competencias específico.

A esto se refiere el cambio conceptual que está detrás de sistemas de evaluación de docentes como el del NBPTS, recordando que las ideas sintetizadas por Lee Shulman en un famoso artículo (1987b) sobre la base de conocimiento que sustenta la enseñanza fueron esenciales para reformular el discurso sobre la manera en que pensamos el trabajo docente y sobre la forma posible de evaluarlo.

Cuando el propósito de la educación básica para la mayoría de los estudiantes era sólo que aprendieran a leer, escribir y contar en forma elemental, a memorizar datos básicos de historia y geografía, a respetar normas sociales fundamentales y poco más, el aprendizaje y la enseñanza eran relativamente simples, al igual que la evaluación de uno y otra. El trabajo de los docentes no debía ir más allá de exposiciones repetitivas, el de los estudiantes de repasos repetitivos también, y la evaluación podía consistir en preguntas orales o escritas que tenían una sola respuesta correcta simple. No es sorprendente que los investigadores educativos encontraran que en las aulas de la segunda mitad del siglo XX seguía presente, en versión laica, la vieja enseñanza tipo catecismo (Resnick *et al.*, 2010: 398-400).

Para que un grupo haga una y otra vez ejercicios rutinarios y logre aprendizajes básicos sí es suficiente un maestro no profesional, que tenga conocimientos básicos y un mínimo de regularidad y orden en su trabajo.

Pero eso no basta para que todo alumno, incluidos los que tienen más dificultades, desarrollen competencias cognitivas y no cognitivas complejas, para hacer frente a los retos de una economía competitiva y una democracia madura, en el siglo XXI. El reto para aprendices y docentes es mayúsculo. Aprender a razonar, inferir, analizar y sintetizar, juzgar críticamente, ser creativo, trabajar en equipo, respetar a los demás y hacerse respetar a sí mismo, etcétera, es difícil para los alumnos, como lo es para el docente desarrollar tales competencias, y evaluar que todo ello ocurra.

Un cuestionario con preguntas de respuesta estructurada permite saber si alguien comprende básicamente lo que lee, si hace operaciones matemáticas sin errores o si recuerda datos básicos, pero no es muy adecuado para valorar competencias más complejas. Una lista de cotejo para usar en una observación permite verificar si un docente lleva a cabo o no, muchas o pocas veces, conductas precisas, pero difícilmente permitirá apreciar el sustento y la consistencia en la aplicación de una estrategia docente compleja con un grupo de alumnos de características precisas.

La búsqueda de acercamientos más poderosos se entiende precisamente por la necesidad de no limitarse a evaluar procesos simples de aprendizaje o de enseñanza, sino de incluir todas las dimensiones del perfil de un buen docente, en el contexto de propósitos educativos ambiciosos.

Para desarrollar una buena evaluación de la docencia hay que abandonar la idea de que la buena enseñanza se identifica con un estilo en particular; un estilo tradicional de enseñanza no necesariamente es inadecuado, hay maestros que pueden dar buenas clases de tipo expositivo, y no porque en un aula veamos gran actividad y muchos materiales se garantiza que hay un buen nivel de aprendizaje.

Según Haertel, cuando se privilegia un enfoque pedagógico particular

puede tratarse de un modelo de enseñanza muy efectivo para alcanzar ciertos objetivos de aprendizaje con algunos alumnos, pero se trivializa el trabajo docente si se afirma que toda la enseñanza en aula, en cualquier grado o área curricular, y respecto a cualquier objetivo educativo, se debe llevar a cabo de la misma manera. Los mejores maestros dominan un amplio repertorio de enfoques instruccionales y varían su enseñanza en formas complejas (Shulman, 1987b). Las prescripciones simplistas de lo que sería "una enseñanza efectiva" usurpan el papel profesional y la responsabilidad del maestro (Haertel, 1991: 4).

La dificultad de poner en práctica estrategias de enseñanza sofisticadas para desarrollar competencias complejas está presente no sólo en las escuelas de países como México sino también en las de otros altamente industrializados.

La investigación sobre las prácticas de evaluación formativa en aula, por ejemplo, muestra que en México muchos docentes dicen hacer evaluación formativa, pero las ideas que tienen al respecto son superficiales y su práctica no coincide con ellas (Martínez Rizo y Mercado Salas, en prensa), lo mismo ocurre en otros países de América Latina (Ravela, Picaroni y Loureiro, 2009), y también en los Estados Unidos (Stiggins y Conklin, 1992).

El punto de partida del trabajo del NBPTS fue, precisamente, reconocer que las formas habituales de trabajo en aula distaban de corresponder a las exigencias de sociedades que piden a los ciudadanos el dominio de competencias complejas:

Aulas auto-contenidas, trabajo con todo el grupo, instrucción centrada en el libro de texto, enseñanza como clase frontal, aprendizaje como adquisición pasiva de hechos, pruebas estandarizadas, todos esos patrones del trabajo escolar son tan familiares como el polvo del gis. Constituyen un currículo nacional no explícito que, como una dieta rígida, no responde adecuadamente a las necesidades educativas de una sociedad diversa y dinámica. Los buenos maestros, desde luego, se apartan de muchas maneras de esas rutinas (NBPTS, 1989, citado por Haertel, 1991: 9).

Las consecuencias para la evaluación de docentes son claras:

La evaluación de maestros profesionales deberá diseñarse para asegurar altos estándares de práctica profesional, no para producir respuestas aprobadas a preguntas de opción múltiple que trivializan las complejidades de la enseñanza, o para monitorear la conformidad con una lista de cotejo de conductas de aula aprobadas también (Haertel, 1991: 7).

b) La dificultad de la actualización de los maestros en servicio

De lo anterior se deriva una consecuencia importante, que amplía y precisa la idea de que no bastará hacer buenas evaluaciones de docentes para que mejore la calidad de su trabajo, y para que se puedan esperar mejoras similares en lo que se refiere a los resultados de aprendizaje de los estudiantes: transitar de una forma convencional de enseñar a una docencia profesional rica no es fácil, ya que se deben superar muchos obstáculos (*cf.* Martínez Rizo, 2013).

Por ello, para conseguir que muchos maestros den el paso no bastan actividades de desarrollo profesional de corta duración y escasa consistencia, como las que abundan en muchos sistemas educativos, como expresa Fullan a principios de este siglo:

Si bien la inversión en actividades de actualización profesional para maestros ha sido durante muchos años un rasgo del panorama educativo, la evidencia de que producen alguna diferencia en el desempeño de los alumnos es deprimentemente escasa. Nada ha prometido tanto y ha sido un desperdicio tan frustrante como los miles de talleres y conferencias que no trajeron cambio significativo en la práctica una vez que los maestros regresaron a sus aulas (según Wiliam, 2007: 187).

Para modificar significativamente prácticas profundamente arraigadas serán necesarios programas de actualización prolongados, sustentados teóricamente, y basados en el trabajo de comunidades de aprendizaje de maestros dispuestos a hacer el esfuerzo correspondiente, y que cuenten con apoyo efectivo por parte de las instancias superiores a la escuela (*cf.* Martínez Rizo, 2012c). ●

Conclusiones

Durante los siglos XIX y XX, y lo que va del XXI, la evaluación de los maestros se ha mantenido como un tema de indudable importancia para la calidad educativa, pero no ha sido resuelto de manera satisfactoria.

En los países europeos y sus colonias, en especial las de Francia y España, la evaluación de los docentes se basó tradicionalmente en el trabajo de inspectores, pues se reconocía la necesidad de aprovechar el juicio de expertos que tuvieran contacto directo con maestros y escuelas. En los Estados Unidos y otros países en los que la investigación educativa se desarrolló con más fuerza, se ha dado más peso al uso de instrumentos estandarizados, de los que se espera que aumenten la confiabilidad de la evaluación al reducir la subjetividad.

El paso de escalas diferenciadas de salarios según características raciales o de género dejó el lugar a escalas únicas basadas en escolaridad y antigüedad, pero la experiencia y la investigación coinciden en mostrar que tampoco esas escalas únicas de tipo escalafonario cumplen la función de distinguir a los mejores maestros de los menos buenos.

En muchos casos la evaluación de los docentes ha tenido serias deficiencias, cuyo resultado se ha reflejado en decisiones arbitrarias por parte de las instancias responsables de la evaluación, mientras los maestros permanecían en un estado de indefensión. En el extremo opuesto, el peso político de los sindicatos de maestros hizo que en otros casos se lograra que las autoridades se limitaran a calificar el trabajo de los maestros como satisfactorio o no, y que la calificación favorable se aplicara en prácticamente todos los casos, exigiendo al docente poco más que asistir a la escuela, mantener el orden y cumplir con las tareas administrativas correspondientes. En sistemas educativos menos sólidos ni siquiera deficiencias serias, ausencias reiteradas, faltas administrativas, e incluso delitos graves han sido motivo de despido. Los sistemas de inspección mismos, igual que los puestos de dirección de escuelas, fueron controlados por los sindicatos, lo que anuló totalmente su función evaluadora.

En cuanto a instrumentos estandarizados, las pruebas simples de conocimientos y los protocolos de observación basados en listas de cotejo igualmente sencillas, sobre la presencia de conductas estereotipadas, tampoco dieron resultados satisfactorios, lo que llevó a la búsqueda de mejores acercamientos.

El uso de resultados de aprendizaje de los alumnos como indicador de la calidad de su maestro es atractivo a primera vista, sobre todo si se usan modelos de valor agregado, y parece posible al contar con pruebas estandarizadas censales, como ocurre hoy en no pocos países. Sin embargo, una consideración cuidadosa de los alcances y límites de esos modelos lleva a la conclusión de que pueden servir para evaluar sistemas o programas que involucren a muchas personas, pero que

no son adecuados para sustentar por sí solos decisiones fuertes sobre maestros o escuelas individuales. No hay duda de que el aprendizaje de los alumnos deberá ser considerado en cualquier buen sistema de evaluación de docentes, pero con otros elementos y siempre que la evaluación de aprendizajes no se reduzca a los aspectos más fáciles de medir con preguntas de respuesta estructurada, sino que incluya competencias cognitivas y no cognitivas de distinto nivel.

La conciencia de la complejidad de la práctica docente llevó a buscar sistemas de indicadores múltiples que combinan pruebas de conocimientos; autorreportes de los docentes, diarios o bitácoras; encuestas que recogen las opiniones de alumnos y padres de familia; observaciones en aula y portafolios y otros acercamientos basados en la revisión de materiales y productos de la docencia.

Cada acercamiento tiene puntos fuertes y débiles; ninguno es superior o inferior a otros en forma absoluta, sino que la calidad de la información que proporciona cada uno depende del aspecto a estudiar y del propósito del estudio, de la calidad del instrumento y de la forma en que se aplique y se analicen sus resultados.

Además hay que destacar la *dimensión formativa* que puede tener la evaluación de maestros, al identificar deficiencias particulares de unas prácticas y fortalezas de otras, para diseñar actividades de actualización y otras formas de intervención que mejoren el trabajo de los maestros menos experimentados o menos eficaces.

CONSIDERACIONES PARA MÉXICO

Los esfuerzos del gobierno federal por hacer obligatoria una evaluación para los maestros de educación básica han provocado una fuerte polémica. En las posturas extremas del debate es posible identificar puntos que coinciden en algún grado con ideas expresadas en estas páginas, a la vez que elementos que se contraponen profundamente a las ideas que se sostienen en ellas.

La preocupación por asegurar que todo alumno sea atendido por un docente calificado es ampliamente compartida, como puede serlo el interés por identificar a quienes destaquen por su dedicación a la docencia, y también por la calidad con la que lleven a cabo su tarea, de modo que su trabajo pueda ser reconocido con estímulos simbólicos y materiales, individuales y colectivos. El que la evaluación sirva para identificar las llamadas "áreas de oportunidad" y, con base en ello, para desarrollar actividades de actualización no parece enfrentar problemas. El punto más conflictivo es el relativo a la posibilidad de que de la evaluación se deriven consecuencias fuertes para los maestros, en términos no sólo de estímulos o ascensos, sino también de sanciones, en particular la posibilidad de que una persona pueda ser despedida por haber obtenido un resultado desfavorable.

Los elementos desarrollados en este documento permiten compartir el punto de vista de quienes consideran inaceptable que un maestro pueda ser despedido con base en sus resultados en una prueba con preguntas de opción múltiple. Al mismo tiempo, no parece difícil estar de acuerdo en que faltas graves como el abandono del trabajo o el abusar de los alumnos debieran sancionarse

con medidas fuertes, incluyendo el despido. Esto último, sin embargo, se concibe como algo distinto de la evaluación. Un documento del Departamento de Investigaciones Educativas, por ejemplo, propone hablar en ese caso de “auditar el cumplimiento del trabajo docente”, con lo que sí se podrían asociar medidas como el despido, en tanto que la evaluación tendría un enfoque básicamente formativo (Weiss, 2013).

La distinción entre evaluación y auditoría parece solamente semántica. El punto clave es la manera en que se defina qué es un buen docente. En este documento se propone que, en el caso de un sistema educativo como el mexicano, esa definición debería incluir aspectos de la llamada “normalidad mínima”, que no parecen necesarios en sistemas educativos más consolidados.

Si se acepta esa idea, la evaluación de un maestro deberá incluir no sólo sus conocimientos de los contenidos a enseñar, de pedagogía en general y pedagógicos del contenido, así como sus prácticas docentes, sino también su asistencia, el cumplimiento de sus obligaciones básicas, el mantenimiento de un ambiente sano y seguro, el cuidado de los alumnos, etcétera.

El siguiente punto del razonamiento es que no todos los aspectos se evalúan de la misma forma ni por los mismos actores. Valorar si un maestro cumple con la normalidad mínima no implica pruebas ni protocolos de observación sofisticados, sino personal directivo que tenga la autoridad y la competencia necesarias. La identificación de fallas graves en estos aspectos debe dar lugar a las sanciones correspondientes, cuidando la objetividad e imparcialidad, así como el respeto de los derechos de los involucrados, lo que no tiene por qué implicar procedimientos tortuosos ni tiempos interminables.

Por lo que se refiere a la evaluación de conocimientos y prácticas de los docentes, la amplia experiencia recogida en este documento lleva a la conclusión clara de que no deberá basarse en instrumentos simples, sino que implicará el uso de varios acercamientos, con características técnicas y condiciones de aplicación (tiempo, personal calificado) que permitan obtener información de buena calidad.

Quienes temen que se establezcan evaluaciones de las que nunca se deriven consecuencias fuertes muchas veces tienen en mente casos extremos de abusos, abandono del empleo y otros. Esas faltas no se deberían aceptar, su corrección no debería llevar mucho tiempo, y hacerlo sería una mejora importante en algunas de las zonas con problemas educativos más graves.

Pero medidas fuertes adoptadas con base en evaluaciones simplistas pueden dar lugar a claras injusticias y además tener resultados contraproducentes para la mejora de la calidad educativa, que es el propósito final que debe guiar el desarrollo de un sistema de evaluación de maestros. Un ejemplo imaginario, pero con sustento real, podrá ilustrar mejor esta idea.

Pensemos en un docente que trabaje en una escuela situada en una comunidad remota, con muchas carencias y que atiende a niños de contexto vulnerable. Imaginemos que se trata de una persona dedicada a su trabajo y preocupada por sus alumnos, que nunca falta a sus clases y que hace todo lo posible por que el grupo a su cargo aprenda lo más posible. Pensemos también que ese docente hizo la primaria en una escuela multigrado, pasó a una telesecundaria y luego a

un telebachillerato, para luego formarse en una normal llena de carencias, y que tiene mala ortografía y no acaba de entender los quebrados. Sus resultados en una prueba de conocimientos son realmente bajos. ¿Deberíamos despedirlo sin más, o bien apoyarlo para que se supere? Creo que la segunda opción es la única adecuada en un contexto como el mexicano.

Muchos maestros y maestras del país tienen deficiencias serias, y hay algunos que además no cumplen con sus obligaciones más elementales, pero también hay otros que se parecen al docente imaginario del párrafo anterior.

La evaluación docente contribuirá a que la calidad de la educación mexicana mejore, como todos queremos, si toma en cuenta la situación real de los maestros y las maestras del país; si precisa los elementos de la normalidad mínima que todos deberán cumplir y los procedimientos justos y expeditos para remover a quienes no lo hagan; si desarrolla, con el rigor técnico necesario, instrumentos con los que se obtenga información de buena calidad sobre aspectos complejos de la práctica docente; si establece criterios razonables para reconocer desempeños ejemplares y corregir los deficientes, y, lo que será clave, si es acompañada por estrategias consistentes de apoyo a escuelas y docentes para la mejora.

Citando a un ex ministro de Educación de Chile:

una recomendación para quienes estén verdaderamente interesados en acometer cambios significativos y perdurables en educación: deben evaluar con rigor los resultados, procesos y organización existentes, sin menospreciar o desechar *a priori*. Hay mucha gente comprometida de corazón, que efectúa loables esfuerzos y se desempeña con excelencia. La tentación de hacer juicios descalificatorios a la bandada acarrea desánimo a los mejores. Y, en educación, su buena disposición y entusiasmos son fundamentales. Se debe actuar de manera firme con los negligentes y estimulante con los diligentes. La escasa comprensión de la complejidad de los factores involucrados en el mejoramiento de la educación conduce a formular apreciaciones tajantes como que “la educación es un desastre”, que no sólo menosprecian los logros obtenidos y suelen venir acompañadas de recetas fáciles e inductoras, sino que alejan cualquier posibilidad de solución y desmoralizan a profesores, directivos, expertos, funcionarios, padres y estudiantes (Bitar, 2011). ●

Apéndice

La evaluación de los docentes en Iberoamérica

Los países iberoamericanos cuentan con sistemas de tipo escalafonario para definir el salario de los maestros, en su calidad de funcionarios públicos, y también con inspectorados, dentro de la tradición napoleónica y española. La operación de esos sistemas es muy desigual. Como ha ocurrido en otros lugares, en épocas recientes en varios países de la región han surgido nuevos modelos de evaluación de docentes, que buscan la forma de diferenciar el salario según el desempeño de los maestros y consideran la posibilidad de que decisiones fuertes como despedir a un docente se hagan con base en los resultados de las evaluaciones.

En este apéndice se presenta una visión panorámica de la situación que prevalece en varios países de Iberoamérica en cuanto a evaluación de docentes, a partir de un trabajo para la Organización de Estados Iberoamericanos para la Educación, la Ciencia y la Cultura (OEI). Se presentan primero los casos de España y Portugal, los dos países europeos de la región cuyo sistema educativo se puede considerar más consolidado, y que, en el primer caso, se apega más al modelo tradicional, y, en el segundo, ha avanzado más en la introducción de innovaciones. Luego se describe la situación de una docena de países de América Latina, comenzando también con los que presentan situaciones más tradicionales para seguir con los que han introducido innovaciones, con grados muy diferentes de avance.

ESPAÑA

Sigue basando el salario de docentes no universitarios en escalafones definidos en la normatividad y en un inspectorado de fuerte tradición. En las 17 comunidades autónomas se debe tener título de maestro y pasar una oposición para ser docente de escuela pública; en el nivel de secundaria, además del título de licenciado o grado se debe hacer un máster y aprobar un concurso; en educación infantil y primaria el maestro puede ser director, jefe de estudios o secretario del centro escolar en que trabaje; en secundaria puede llegar a catedrático y ser jefe de departamento.

El nombramiento del director de una escuela corresponde a una comisión *ad hoc* que analiza el currículo del candidato y su proyecto de dirección. El acceso al cuerpo de catedráticos se hace también por oposición, considerando antigüedad, formación y puestos desempeñados.

Después de ganar una plaza, por lo general el profesorado no debe someterse a evaluaciones periódicas. El nombramiento de director de escuela se debe renovar cada cuatro años. Una evaluación positiva implica mayor salario; después

de tres evaluaciones favorables un director se puede presentar a una oposición restringida para formar parte del Cuerpo de Inspectores. Los docentes sólo pueden ser removidos mediante un expediente disciplinario por una falta grave o muy grave.

PORTUGAL

Apartir de un sistema tradicional ha introducido innovaciones para dar retribuciones diferenciadas. Los docentes pueden acceder a varios niveles de salario, según evaluaciones de su desempeño en las que participan el presidente del consejo general de la escuela (formado por docentes, padres de familia, alumnos y otros); el director; el consejo pedagógico y su sección de evaluación docente, así como evaluadores externos e internos. Se toman en cuenta tres dimensiones de experiencia y preparación:

- Dimensión científica y pedagógica con 60% de peso, que sube a 70% si hay evaluación externa basada en observación del trabajo en aula.
- Participación en la escuela y la relación con la comunidad, 20%.
- Formación continua y desarrollo profesional, 20%.

La evaluación externa con observación en aula del trabajo del docente puede dar lugar a una remoción después de dos ciclos seguidos de evaluación negativa. Un primer resultado negativo trae como consecuencia que deja de contarse el tiempo de servicio del ciclo de evaluación de que se trate para efectos de ascenso en la carrera docente, y la obligación de llevar y terminar con éxito un programa de formación de un año de duración propuesto por los evaluadores y aprobado por el consejo pedagógico, que incluye la observación en el aula. Un segundo resultado insuficiente implica el inicio de un proceso de averiguación que puede culminar en la remoción. Para los docentes con nombramiento no definitivo dos resultados negativos consecutivos implican que no podrán ser admitidos a cualquier concurso para reclutar personal docente en los tres años escolares siguientes.

Para ingreso, promoción y permanencia en puestos docentes, la primera instancia es el director de la escuela; la segunda es la Dirección General de Administración Escolar (antes Direcciones Regionales de Educación), y la tercera y última es el Ministerio de Educación y Ciencia. Los sindicatos de profesores son escuchados y negocian con el Ministerio la definición de los modelos de evaluación.

Por lo que se refiere a los países latinoamericanos, hay sistemas tradicionales cuyo funcionamiento parece más o menos consolidado.

URUGUAY

Cuenta con una normatividad precisa y actualizada, el *Estatuto del Funcionario Docente*, cuya última actualización se hizo en 2008 y define tres categorías básicas (docente efectivo, interino y suplente) y un *sistema escalafonario e inspectivo* na-

cional que comprende cinco subescalafones especiales para educación primaria, secundaria, técnico-profesional, de formación y perfeccionamiento docente, y de educación de adultos.

El sistema escalafonario comprende siete grados o niveles a los que se accede con base en la puntuación que se obtenga en cuanto a aptitud docente (hasta 100 puntos), antigüedad (20) y "actividad computada" (20); además deben aprobarse los cursos que se reglamenten para cada cargo.

La asignación de puestos de trabajo y los ascensos se hacen por concursos ante un tribunal. Para determinar entre un grupo de aspirantes quién accederá a un puesto determinado, se considera en primer lugar a quienes tengan el grado más alto del escalafón; entre los que tengan el mismo grado se preferirá a los de mayor antigüedad, y luego a los de mayor valor en el rubro de aptitud docente.

Una Junta Calificadora evalúa la aptitud docente, con base en los informes de las visitas anuales de inspección al aula, en el juicio del director de la escuela, el cumplimiento de registros y documentación, los cursos de capacitación y perfeccionamiento, actividades de investigación u otras relacionadas con la docencia, y en observaciones de orden disciplinario.

Los juicios de los inspectores deben considerar la capacidad técnico-pedagógica; la conducción del proceso de enseñanza aprendizaje; la planeación y el desarrollo del curso; los aprendizajes alcanzados por los alumnos y su capacidad para seguir aprendiendo; el clima de trabajo; el respeto a los alumnos, así como la promoción de su capacidad de autodeterminación y trabajo creativo. El juicio del director de la escuela tendrá en cuenta la aptitud y preparación del maestro, su iniciativa, su disposición al trabajo colaborativo, su asiduidad y puntualidad, las relaciones humanas, el interés por los alumnos y su trato con ellos, sus trabajos de investigación, la contribución a la formación de futuros docentes, la integración de tribunales y la colaboración para la inserción de la escuela en su medio social.

El valor mínimo aceptable para el rubro de aptitud docente es el de 51 puntos de los 100 posibles; si un docente recibe una calificación menor su caso pasa a estudio de la Junta de Inspectores, que es la instancia facultada para hacer una declaración de ineptitud, que puede dar lugar al cese. Fuera de los casos de conducta grave, omisión o delito; abandono del cargo, o ineptitud física o mental acreditada por junta médica, el cese de un maestro sólo puede proceder por este tipo de declaración comprobada para el ejercicio de la docencia.

CUBA

Un sistema sólido de evaluación de docentes basado en el mecanismo de la inspección. La evaluación puede dar lugar a la promoción del interesado a cargos de mayor responsabilidad, como la dirección del centro escolar, o a una remuneración adicional.

Todos los meses, los directores de centros, de común acuerdo con la organización sindical, evalúan sistemáticamente a los docentes analizando el cumplimiento de los objetivos, las funciones y el plan de trabajo individual, lo que permite

puntualizar los aspectos deficientes que se deben mejorar. El proceso concluye, en dos momentos del curso escolar, con un "certificado de evaluación".

Existe la posibilidad de remover a un docente si de manera consecutiva tiene un desempeño negativo y no logra resultados aceptables en el aprendizaje, o bien por incumplimiento de sus obligaciones, o si ha incurrido en faltas graves de la disciplina laboral o en el ejercicio de la docencia. En esos casos se da por concluida la relación laboral o se cambia a la persona a cargos menos complejos. Se toman en cuenta:

- Los resultados de la evaluación del desempeño del docente, que valora disciplina del trabajo, calidad de la docencia, fundamentalmente en clase y aprendizaje de los educandos.
- El vínculo de trabajo con organizaciones estudiantiles, políticas y de masas, así como con la comunidad donde residen los estudiantes.
- La contribución a la formación política ideológica, fundamentalmente en lo referido a la formación de valores, al dominio de la Historia de Cuba y su vínculo con la historia local.
- Los resultados del trabajo educativo considerando los indicadores de eficiencia educacional: asistencia, retención, resultados de las comprobaciones de conocimientos y habilidades, grado de desarrollo de las capacidades, y hábitos en la formación de convicciones de los estudiantes.
- La atención a las diferencias individuales y su contribución al trabajo correctivo; el cumplimiento del reglamento escolar y el cuidado de la propiedad social e individual en el trabajo productivo y socialmente útil, en las actividades de estudio y trabajo, en el trabajo de orientación profesional y vocacional y en la labor sistemática con la familia.

GUATEMALA

La Ley de Dignificación y Catalogación del Magisterio Nacional define seis clases a las que pueden acceder los maestros. A la clase "A" le corresponde el sueldo básico, y a las cinco clases siguientes corresponde un incremento de 25% en cada caso, de manera que la clase "B" tiene un ingreso 25% superior al sueldo básico; la clase "C", 50% más; la clase "D", 75%; la "E", 100%, y la clase "F" 125% sobre el sueldo básico.

Para acceder a esas clases se utiliza un sistema de puntuación que considera ante todo el tiempo de Servicio (40% del total), pero también la calidad (20%), la superación (20%), los méritos especiales (5%) y los servicios extra cargo (15%). Los puntos se comprueban anualmente y se totalizan cada cuatro años para cada docente, por una Junta Calificadora de Personal que valora la superación profesional, la calidad de los servicios y los méritos obtenidos con base en una Hoja de Servicio que llenan los directores de establecimiento, y comprende lo siguiente:

No se exige a los docentes someterse a algún tipo de evaluación después de cierto tiempo para permanecer en el puesto, ni existe la posibilidad de que un docente sea removido de su puesto como resultado de alguna evaluación.

El proceso para ingresar al servicio es de tipo oposición, con jurados en los niveles municipal, departamental y nacional. El Sindicato de Trabajadores de la Educación de Guatemala tiene un peso significativo en las decisiones, dada su participación en las mesas técnicas que forman parte de la Junta Calificadora de Personal y de los Jurados de Oposición en toda la estructura.

NICARAGUA

La Ley de Carrera Docente señala que el ingreso a la docencia se define con base en la preparación profesional y la experiencia de los aspirantes, pero añade que donde no haya recursos humanos calificados para atender la demanda se puede contratar personal no titulado que posteriormente se incorpore a un programa de profesionalización docente.

Para subir de nivel se toman en cuenta la experiencia; la preparación académica; las capacitaciones técnicas, metodológicas y administrativas, y el contar con un expediente libre de amonestaciones. Los títulos que evidencian la preparación académica son el elemento más importante, seguido por el análisis del archivo docente que contiene las evidencias de desempeño laboral (planes, cuadernos de registro, etcétera) y los informes de visitas de asesoría recibidas en el aula. También son importantes los reconocimientos cuando el docente ha asumido responsabilidades, su relación con los padres y su integración en la comunidad.

No se precisan los tiempos para realizar los procesos de evaluación. Los criterios establecidos por la Ley son:

- Capacidad y eficiencia en el trabajo, que incluye:
 - Dominio cognoscitivo de la ciencia o técnica y pedagogía.
 - Capacidad y habilidad para transferir y aplicar el conocimiento.
 - Dirección técnica y pedagógica del proceso de enseñanza-aprendizaje, planificación, organización, ejecución y evaluación.
 - Cooperación con actividades del Ministerio, el centro y la comunidad.
- Ética profesional.
- Relaciones humanas.

Puede haber procesos de formación ligados a una evaluación, para ayudar a los docentes a mejorar su desempeño, con apoyo en la asesoría pedagógica.

Existe la posibilidad de remoción, pero una evaluación no es suficiente para remover a alguien de su cargo. La Ley de Carrera Docente protege la estabilidad laboral. Sólo procederá el retiro o la suspensión temporal del servicio por abandono del cargo o incumplimiento reiterado e injustificado de las funciones y obligaciones propias del mismo, o por haber cometido delitos y haber sido condenado por sentencia firme.

La remoción producto de evaluación solamente ocurre si existen evidencias de incumplimiento o deficiencia laboral demostrada en varios informes de seguimiento para finalmente tomar la decisión. El reglamento de la Ley de Carrera Docente establece: "Los resultados de la evaluación en principio tendrán carácter

educativo y retroalimentador, pero también se utilizarán para retiro y otros movimientos de personal”.

HONDURAS

Los elementos que definen el salario son los tradicionales: escolaridad formal, escolaridad adicional y antigüedad, con el mismo peso. Después de que alguien ingresa al servicio sólo se le puede suspender temporal o definitivamente cuando hay causas muy graves. No hay procesos de evaluación del desempeño.

Hay seis organizaciones sindicales que defienden a sus agremiados con mucho poder social y económico. Hay procesos de capacitación o actualización docente como respuesta a los resultados de la evaluación del rendimiento académico de los estudiantes, a los que se aplican pruebas para ello, pero no a partir de evaluaciones de los docentes, que no existen.

La nueva Ley Fundamental de Educación considera que a partir de 2014 se realice una evaluación del desempeño en la que participen todos los docentes en igualdad de condiciones, con cuyos resultados se hagan acreedores a incentivos temporales.

EL SALVADOR

La Ley de la Carrera Docente establece un sistema de incrementos salariales por antigüedad: después de los primeros cinco años de trabajo, un aumento de 10%; después de 10 años, 8%; luego el aumento es de 6% cada cinco años hasta los 35. La Ley clasifica también a los maestros en Docentes Nivel 1 y Nivel 2, según su formación académica. La instancia que asigna esos niveles es el Tribunal Calificador de la Carrera Docente, que considera la escolaridad formal y los cursos de actualización. No existen evaluaciones de desempeño. La posibilidad de que un docente sea removido de su puesto sólo puede ocurrir por las faltas tipificadas en las leyes con sus correspondientes sanciones.

REPÚBLICA DOMINICANA

Un docente puede ascender a puestos de responsabilidad y percibir incentivos que son un porcentaje del salario base: por titulación de 10% a 40%; por años en servicio de 10% a 15%; por evaluación del desempeño de 7% a 35%. El Reglamento del Estatuto Docente define la figura del Educador-Coordenador, una categoría de maestro con mayor responsabilidad docente, pero en la práctica esta figura no existe.

Para ascender a niveles superiores hay que presentar un concurso de oposición. Los aspirantes deben tener la formación y la calificación para el puesto vacante. Hasta ahora sólo hay concursos para ingreso a plazas de maestro y director. La evaluación de desempeño da lugar a recomendaciones de capacitación para los maestros y técnicos.

La Ordenanza 6-2010 reglamenta los concursos para selección de docentes, y establece una distribución de 100 puntos como sigue:

- Perfil profesional (formación y experiencia): 10 puntos
- Prueba escrita: 50 puntos
- Práctica de planificación: 25 puntos
- Entrevista oral: 15 puntos
- Prueba psicológica: cumple o no cumple

Para los puestos en que no hay concursos, los elementos que se toman en cuenta incluyen grado universitario, especialidad en la materia de su competencia, cinco años o más de experiencia y una historial laboral sin registro de faltas e indisciplina. No se establece un peso específico para cada uno de estos aspectos.

Un docente puede ser removido del empleo si incurre en las causas indicadas en el Régimen Disciplinario del Reglamento de Estatuto Docente y la Ley de Función Pública. En la práctica se llega a acuerdos entre la autoridad y el Sindicato, que es parte de la comisión de evaluación de desempeño. No se considera la opción de remover al docente por resultados de la evaluación, y ésta contempla la opción de mejora continua a través de programas de capacitación. Actualmente no hay certificación y recertificación, pero está diseñándose una metodología y los procesos para tal efecto, como exigencia de la Estrategia Nacional de Desarrollo. En 2013 se realizó un diagnóstico para crear el sistema de certificación docente que entró en vigencia gradualmente a partir del 2014.

ECUADOR

El Plan Decenal de Educación 2006-2015 incluye la propuesta de un sistema nacional de evaluación y rendición social de cuentas, que evalúe la gestión del Ministerio, el desempeño de los docentes, el de los estudiantes y el currículo nacional. La evaluación se plantea como un medio para el mejoramiento de los desempeños y procesos, y que no es punitivo ni busca la sanción de los involucrados, sino más bien persigue estimular el esfuerzo realizado por todos los actores educativos y la institución educativa.

La evaluación comenzó en 2008 con una convocatoria voluntaria a todos los docentes del sector público. De unos 200 mil sólo se presentaron 1 569, por lo que se optó por hacer obligatoria la evaluación mediante un acuerdo presidencial de 2009, que oficializó el sistema de evaluación y rendición social de cuentas, al tiempo que estableció el mecanismo para la evaluación de los docentes. La evaluación obligatoria fue aplicada a una muestra de 25% de los maestros de todo el país, y se planeó continuar evaluando 25% cada año hasta culminar en el 2013. La evaluación tiene un componente interno y uno externo, cada uno con 50% del peso total:

La evaluación interna comprende una autoevaluación (5% de la nota); una coevaluación por un colega de la misma área (5%); una por parte de los directivos (5%);

otra por los padres y madres de familia (8%); una más por los estudiantes (12%), y una observación del aula (15%).

La externa es una prueba escrita que mide conocimiento específico (30%), pedagógico (10%), y comprensión lectora (10%). Para los docentes de planteles bilingües se propone una prueba de lengua materna. En cuanto a conocimientos específicos hay pruebas distintas para docentes del primer año de educación básica y docentes de segundo a séptimo año de ese nivel. La escala de valoración va de 1 a 100, siendo 90 o más excelente; de 80 a 89 muy bueno; de 60 a 79 bueno; y menos de 60 insatisfactorio.

Se ofrecen estímulos económicos y desarrollo profesional a los docentes con resultados excelentes. Los que obtengan una calificación insatisfactoria tendrán una oportunidad adicional para someterse a una nueva evaluación, y, en caso de no aprobarla, podrían ser separados de su puesto.

Ante la resistencia de los docentes para presentar las pruebas, el gobierno amenazó con destituir al que se negase a ello. El Decreto Ejecutivo 1740 obliga a los docentes a ser evaluados; en caso contrario serían removidos y declarados "incompetentes profesionales". En las evaluaciones de 2008 a 2012, de 60 073 docentes evaluados, 1.18% obtuvo una calificación de excelente; 34.41%, muy bueno; 61.89%, bueno, y 2.52%, insatisfactorio

En marzo de 2011 se aprobó la Ley Orgánica de Educación Intercultural, que dio un giro a la evaluación docente. Al Ministerio de Educación toca ahora formular los estándares de calidad para el desempeño profesional de docentes y directivos. Para valorar su cumplimiento, en noviembre de 2012 se creó el Instituto Nacional de Evaluación Educativa (INEVAL).

BRASIL

Por la organización federal del país, los estados y municipios son responsables de regular sus propios sistemas de enseñanza, pero hay una red federal de docentes en educación básica, técnica y tecnológica. Para ellos, y además de la posibilidad de acceder a puestos de dirección, la Ley 12 772 de 2012 establece que la carrera del magisterio comprende cinco clases con remuneraciones ascendentes: Profesor Asociado 1, Profesor Asociado 2, Profesor Asociado 3, Profesor Asociado 4 y Profesor Titular.

La Ley 12 772 establece normas para avance y promoción en la carrera, que todavía deben ser reglamentadas, a partir de dos criterios generales: tener 24 meses de actuación efectiva en cada nivel y aprobar la evaluación de desempeño individual. Para pasar de una clase a la siguiente se requerirá obtener aprobación en un proceso de evaluación de desempeño. Para acceder a la clase de Profesor Titular, además de aprobación en la evaluación de desempeño, se requerirá tener título de doctor y obtener aprobación de un memorial que considere las actividades de enseñanza, investigación, extensión, gestión académica y producción profesional relevante, o defensa de una tesis académica. El sistema de enseñanza federal no establece pesos para los distintos criterios de promoción y progresión en la carrera.

Los profesores de la red federal están sometidos al régimen jurídico de los servidores públicos (Ley 8 112/90), que establece la obligatoriedad de un periodo de prueba (*estágio probatório*) con duración de tres años, en el cual son evaluados anualmente por su jefe inmediato considerando aptitud y capacidad para el cargo, con énfasis en asiduidad, disciplina, capacidad de iniciativa, productividad y responsabilidad.

Cumplido y aprobado ese periodo, el servidor es “confirmado” en el puesto adquiriendo estabilidad, y sólo puede ser retirado por falta grave. No se permite el despido con base en evaluaciones no satisfactorias; las leyes de Brasil no permiten la remoción de un servidor público a título de castigo.

En algunos estados hay evaluaciones para asignar estímulos económicos o salarios diferenciados, como Minas Gerais y São Paulo. En esos estados y en Amazonas, Ceará, Espírito Santo, Pernambuco y Río de Janeiro se utilizan los resultados de los estudiantes en pruebas estatales, junto con otros indicadores, para definir esos estímulos (Morduchowicz, 2011).

PERÚ

Las modificaciones más recientes son las de la Ley de Reforma Magisterial de noviembre de 2012, que establece evaluaciones que permiten ascender en ocho niveles o escalas, las cuales implican una remuneración adicional que va de 100% a 260%, o el acceso a cargos de más responsabilidad. Un comité de evaluación de ascenso toma las decisiones, considerando los conocimientos disciplinares del área y el nivel en que enseña el docente; su dominio de la teoría pedagógica; la calidad de su desempeño laboral; su formación profesional, así como sus “méritos”. El aspecto de más peso es el desempeño laboral, con base en los resultados de la última evaluación que haya pasado el profesor. Esta evaluación se basa en los criterios del buen desempeño docente contenidos en las políticas del Ministerio de Educación, e incluye necesariamente la evaluación del progreso de los alumnos.

Cada tres años el docente debe someterse a una evaluación basada en esos criterios, con la posibilidad de remoción si su desempeño es muy deficiente o si no presenta la evaluación sin causa justificada. Sin embargo, la remoción no puede tener lugar de inmediato: el docente cuyo desempeño sea valorado negativamente debe participar en un programa de superación profesional durante seis meses, tras lo cual debe someterse a una segunda evaluación. Si el resultado vuelve a ser desfavorable tiene una segunda oportunidad de participar en el programa de superación, tras lo cual se le podrá remover si los resultados negativos persisten.

COLOMBIA

El estatuto docente establece las condiciones de ingreso, ejercicio, estabilidad, ascenso y retiro. El ingreso al servicio estatal es mediante un concurso de méritos convocado por la entidad territorial certificada por la Comisión Nacional del Servicio Civil (CNSC). Los aspirantes que cumplen los requisitos fijados presentan pruebas de selección, que valoran aptitudes y competencias básicas; con los

resultados se forma un listado de elegibles, que pasan una prueba psicotécnica, una entrevista y una valoración de antecedentes. Según el mérito de cada elegible se le asigna un lugar dentro del listado para cada puesto. Los resultados se publican en la forma de un listado de elegibles por nivel educativo y área de conocimiento, en orden descendente del puntaje obtenido. Para los cargos directivos la Ley 715 de 2011 establece criterios y procedimientos particulares.

El ejercicio de la carrera docente está unido a la evaluación permanente que verifica el desempeño de las funciones; los niveles de idoneidad, calidad y eficiencia que justifican la permanencia; los ascensos en el escalafón, y las reubicaciones en los niveles salariales. En el pasado, incluso en el régimen definido en 1979, no había una obligación de evaluación permanente de docentes y directivos. Desde 2002, a partir de la publicación del Decreto de Ley 1278, hay tres tipos de evaluación: de periodo de prueba, ordinaria periódica de desempeño anual y de competencias.

Tras el ingreso al servicio, al fin del año escolar un docente tiene que ser evaluado *en periodo de prueba*, en desempeño y en competencias específicas.

Superado el periodo de prueba, docentes y directivos deben sujetarse anualmente a evaluación de desempeño (decreto 3 782 de 2007).

La evaluación de competencias se hará cada vez que la entidad territorial lo considere conveniente, pero sin que en ningún caso transcurra un término superior a seis años, y tendrá carácter voluntario para quienes pretendan ascender en el escalafón. Se hará por grados y por cargos directivos. Debe permitir la valoración de competencias de logro y acción; de ayuda y servicio; de influencia; de liderazgo y dirección; de cognición, y de eficacia personal.

Tanto la evaluación en periodo de prueba como las anuales de desempeño laboral y las periódicas de competencias pueden dar lugar a la separación del servicio. El docente con calificación no satisfactoria (60% o menos) en dos años consecutivos en evaluación de desempeño será excluido del escalafón y retirado del servicio. Los directivos docentes que obtengan una calificación inferior a 60% durante dos años consecutivos serán regresados a la docencia una vez que exista vacante, si provenían del servicio estatal, en cuyo caso percibirán el salario que corresponda a dicho cargo, de acuerdo con el grado y el nivel salarial que tenían. Si no provenían de la docencia estatal serán excluidos del escalafón y retirados del servicio.

CHILE

Es un ejemplo de evaluación de docentes basada en investigaciones rigurosas, que aprovechó la experiencia en evaluación de aprendizajes derivada del Sistema de Medición de la Calidad Educativa (SIMCE); el país cuenta con dos sistemas de estímulos:

- La Asignación de Excelencia Pedagógica (AEP), un programa voluntario para reconocer el mérito de maestros de primaria y secundaria que se sometan a una evaluación basada en un *portafolio* con planes de clases, grabación en video de una lección y prueba de conocimientos disciplinarios y pedagógicos.

El maestro bien evaluado recibe un estímulo equivalente a un mes adicional de salario durante 10 años a partir del nivel básico, que depende de la antigüedad en el servicio.

- El Sistema de Evaluación del Desempeño Profesional Docente (SEDPD), con evaluaciones obligatorias cada cuatro años para verificar el cumplimiento de los estándares del Marco de la Buena Enseñanza, que precisa indicadores y niveles de desempeño mediante rúbricas. La información se obtiene de una autoevaluación (10% del peso total); un informe del director y el jefe técnico pedagógico de la escuela (10%); la evaluación de un docente de otra escuela (20%), y un portafolio con evidencias que incluyen la grabación en video de una clase (60%). Los resultados ubican a cada maestro en un nivel de competencia: insatisfactoria, básica, competente y destacada. Un maestro con resultado insatisfactorio debe seguir un curso de capacitación y volver a someterse a evaluación. Un segundo resultado insatisfactorio implica dejar temporalmente la escuela para dedicarse íntegramente a un programa de desarrollo profesional. Un tercer resultado similar lleva a la separación definitiva del servicio. En cambio, obtener un resultado de competente o destacado permite al interesado presentar un examen sobre la(s) materia(s) que enseña para obtener la Asignación Variable por Desempeño Individual (AVDI), que equivale a un incremento de 5% a 25% de la Remuneración Básica Mínima Nacional. El estímulo de la AVDI es acumulable con el de la AEP (cfr. Morduchowicz, 2011; Manzi, González y Sun, 2011; Vaillant, 2012). ●

Referencias bibliográficas

- Ball, D.L., y Hill, H.C. (2009). Measuring Teacher Quality in Practice. En D. Gitomer (Ed.), *Measurement Issues and Assessment of Teaching Quality* (pp. 80-98). Thousand Oakes: Sage.
- Barrera, I., y Myers R. (2011). *Estándares y evaluación docente en México: El estado del debate*. Santiago-Washington: PREAL. Doc. 49.
- Bitar, S. (2011). Cómo construir consensos en educación. En C. Larroulet y P. Montt, *El acuerdo en educación: Lecciones para el futuro*. Santiago de Chile: U. Diego Portales e Instituto Libertad y Desarrollo.
- Blank, R. K., Porter, A., y Smithson J. (2001). *New Tools for Analyzing Teaching, Curriculum & Standards in Mathematics & Science. Results from Survey of Enacted Curriculum Project*. Washington: Council of Chief State School Off.
- Blodgett, J.H. (1893). *Report on Education in the United States at the Eleventh Census, 1890*. Washington: Government Printing Office.
- Bloom, B. (1968). Learning for Mastery. *Evaluation Comment*, 1(2), 1-12.
- Bloom, B., Hastings, J., Madaus, G. F., et al. (1971). *Handbook on Formative and Summative Evaluation of Student Learning*. Nueva York: McGraw-Hill.
- Bolam, R. (1985). School Inspectors and Advisors. En T. Husén y T. N. S. Postlethwaithe (Eds.), *The International Encyclopedia of Education* (vol. 9, pp. 4429-4432). Oxford-Nueva York: Pergamon Press.
- Bolin, F.S., y Panaritis, P. (1992). Searching for a Common Purpose: A Perspective on the History of Supervision. En Carl D. Glickman (ed.), *Supervision in Transition* (pp. 30-43). Alexandria: Association for Supervision & Curriculum Development.
- Borko, H., et al. (2005). Artifact Packages for Characterizing Classroom Practice: A Pilot Study. *Educational Assessment*, 10(2), 73-104.
- Borko, H., Stecher, B., y Kuffner, K. (2007). *Using Artifacts to Characterize Reform Oriented Instruction: The Scoop Notebook and Rating guide* (CSE Technical Report 707). Los Ángeles: UCLA.
- Braun, H. I. (2005). *Using Student Progress to Evaluate Teachers: A Primer on Value-added Models*. Princeton: ETS. Policy Information Perspective.
- Brennan, R. (2013). Commentary on "Validating the Interpretations and Uses of Test Scores". *Journal of Educational Measurement*, 50(1), 74-83.
- Brown, E.G. (2011, 8 de octubre). *Letter to the Members of the California State Senate*. Office of the Governor.
- Brown Center Task Force on Teacher Quality. (2011). *Comparing Teacher Evaluation Systems*. Brookings Institution. Transcripción recuperada el 25 de febrero de 2013, de: en http://www.brookings.edu/~media/events/2011/4/26%20teacher%20evaluation/20110426_teacher_evaluations_transcript.pdf

- CAP-Ed. (2007). An Introduction to the Inspectorate Model. CAP-Ed Symposium *British Solution for California Public Schools?* Center for Applied Policy in Education, UC Davis School of Education.
- Correnti, R., y Martínez Fernández, J.F. (2012). Conceptual, Methodological & Policy Issues in the Study of Teaching: Implications for Improving Instructional Practice at Scale. *Educational Assessment*, 17(2-3), 51-61.
- Darling-Hammond, L. (1989). The Case for a Supervised Teaching Internship. En ETS. *New Directions for Teacher Assessment. Proceedings of the 1988 ETS Invitational Conference* (pp. 99-114). Princeton: autor.
- Darling-Hammond, L. (2008). Standard Setting in Teaching: Changes in Licensing, Certification & Assessment. En V. Richardson (ed.). *Handbook of Research on Teaching* (4a ed., pp. 751-776). Washington: AERA.
- Darling-Hammond, L. (2012). The Right Start. Creating a Strong Foundation for the Teaching Career. *Phi Delta Kappan*, 94(3), 8-13.
- Darling-Hammond, L., y Sclan, E. (1992). Policy and Supervision. En C. E. Glickman (ed.), *Supervision in Transition* (pp. 7-29). Alexandria: Association for Supervision & Curriculum Development.
- Darling-Hammond, L., y Young, P. (2002). Defining "High Qualified Teachers": What Does Scientifically Based Research Actually Tell Us? *Educational Researcher*, 31(9), 13-25.
- Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., y Rothstein, J. (2012). Evaluating Teacher Evaluation. *Phi Delta Kappan*, 94(3), 8-15.
- English, F. (1992). History and Critical Issues of Education Compensation Systems. En L. E. Frase (ed.). *Teacher Compensation and motivation* (pp. 3-25). Lancaster: Technomic Publ.
- Ferguson, R.F. (2008). *The Tripod Project Framework*, 4, 2013-04-02. Recuperado de: http://www.achievementseminars.com/seminar_series_2008_2009/readings/ferguson_tripod_project.pdf
- Ferguson, R.F., y Ramsdell, R. (2011, 3 de mayo). *Tripod Classroom-Level Student Perceptions as Measures of Teaching Effectiveness*.
- Ferguson, R.F. (2012). Can Student Surveys Measure Teaching Quality? *Phi Delta Kappan*, 94(3), 24-28.
- Fullan, M. (1991). *The New Meaning of Educational Change* (2a ed.). Nueva York-Londres: Teachers College Press.
- Gitomer, D. (ed.) (2009). *Measurement Issues and Assessment of Teaching Quality*. Thousand Oakes: Sage.
- Gitomer, D. (2008). Reliability and NBPTS Assessment. En L. Ingvarson y J. Hattie (eds.). *Assessing Teachers for Professional Certification: The First Decade of the NBPTS* (capítulo 9, pp. 231-253). Ámsterdam: Elsevier-JAI.
- Glazerman, S., et al. (2010). *Evaluating Teachers: the Important Role of Value Added*. Washington: Brown Center on Education Policy at Brookings.
- Glenn, C.L. (2006) [ed. original en inglés, 1988]. *El mito de la escuela pública*. Madrid: Encuentro.
- Gobierno Federal (2011). *Acuerdo para la Evaluación Universal de Docentes y Directivos en Servicio de Educación Básica*. México: SEP-SNTE.

- Gobierno Federal (1973). Reglamento de Escalafón de los Trabajadores al Servicio de la Secretaria de Educación Pública.
- Goe, L., Bell, C., y Little, O. (2008). *Approaches to Evaluating Teacher Effectiveness: A Research Synthesis*. Washington: NCCTQ.
- Gonzalbo Aizpuru, P. (2001). *Educación y colonización en la Nueva España 1521-1821*. México: Universidad Pedagógica Nacional.
- Gratz, D.B. (2009). *The Peril and Promise of Performance Pay: Making Education Compensation Work*. Lanham: Rowman & Littlefield Education.
- Haertel, E.H. (1986). The Valid Use of Student Performance Measures for Teacher Evaluation. *Educational Evaluation and Policy Analysis*, 8, 45-60.
- Haertel, E.H. (1991). New Forms of Teacher Assessment. *Review of Research in Education*, 17, 3-29.
- Hill, H., et al. (2008). Mathematical Knowledge for Teaching & Mathematical Quality of Instruction: An Exploratory Study. *Cognition and Instruction*, 26(4), 430-511.
- Hill, H., Schilling, S.G., y Ball, D.L. (2004). Developing Measures of Teachers' Mathematics Knowledge for Teaching. *Elementary School Journal*, 105, 11-30.
- Holmes Group (1986). *Tomorrow's Teachers*. East Lansing: autor.
- Ingvarson, L., y Hattie, J. (eds.) (2008). *Assessing Teachers for Professional Certification: The First Decade of the National Board for Professional Teacher Standards*. Ámsterdam: Elsevier-JAI.
- Kane, M.T. (2013). Validating the Interpretations and Uses of Test Scores. *Journal of Educational Measurement*, 50(1), 1-73.
- Kane, T.J., y Staiger, D.O. (coords.) (2013). *Gathering Feedback for Teaching: Combining High Quality Observations with Student Surveys and Achievement Gains* (MET Res. Paper). Seattle: Bill-Melinda Gates Found.
- Kelly, S. (ed.) (2012). *Assessing Teacher Quality: Understanding Teacher Effects on Instruction and Achievement*. Nueva York: Teachers College Press.
- Koziol, S.M., y Burns, P. (1986). Teachers' Accuracy in Self-reporting about Instructional Practices Using a Focused Self-report Inventory. *Journal of Educational Research*, 79(4), 205-209.
- Ladson-Billings, G. (2005). No Teacher Left Behind: Issues of Equity and Teacher Quality. En C. A. Dwyer. *Measurement and research in the accountability era* (pp. 141-162). Mahwah: Lawrence Erlbaum.
- Ladson-Billings, G. (2009). Opportunity to Teach: Teacher Quality in Context. En D. Gitomer (ed.). *Measurement Issues and Assessment of Teaching Quality* (pp. 206-222). Thousand Oakes: Sage.
- Lazer, S. (2009). Measuring Teacher and Teaching Quality. En D. Gitomer (ed.). *Measurement Issues and Assessment of Teaching Quality* (pp. 150-157). Thousand Oakes: Sage.
- Leckie, G., y Goldstein H. (2009). *The Limitations of Using School League Tables to Inform School Choice*. University of Bristol.
- Lillis, K.M. (1992). *Improving Basic Education: Preconditions for Successful Inspection and Supervision* (Increasing and Improving the Quality of Basic Ed. Monograph 7). París: IIPPE.

- Luna, E., et al. (2012). La evaluación del profesorado de educación básica en México: Políticas, programas e instrumentos. *Revista Iberoamericana de Evaluación Educativa*, 5(2), 231-244.
- Manzi, J., González, R., y Sun, Y. (2011). *La evaluación docente en Chile*. Santiago: MIDE-Universidad Católica.
- Marshall, K. (2012). Let's Cancel the Dog-and-pony Show. *Phi Delta Kappan*, 94(3), 34-38.
- Martínez Rizo, F. (2012a). *La evaluación en el aula* (vol. I. Promesas y desafíos de la evaluación formativa). Aguascalientes: Universidad Autónoma de Aguascalientes.
- Martínez Rizo, F. (2012b). Procedimientos para el estudio de las prácticas docentes. *Revista Electrónica de Investigación y Evaluación Educativa*, 18(1).
- Martínez Rizo, F. (2012c). El reto de cambiar las prácticas de evaluación en aula de maestros de primaria. Consideraciones metodológicas sobre la capacitación a propósito de una experiencia. *Alternativas. Espacio Pedagógico*, 17(66-67), 43-59.
- Martínez Rizo, F. (2013). Las dificultades para implementar la evaluación formativa. *Perfiles Educativos*, 35(139), pp. 128-150.
- Martínez Rizo, F., y Mercado Salas, A. (en prensa). Exámenes aplicados y tareas asignadas en primaria de Nuevo León. En G. Ruiz Cuéllar (ed.). *La evaluación en el aula: Diagnóstico*. Aguascalientes: Universidad Autónoma de Aguascalientes.
- Matsumura, L., y Pascal, J. (2003). *Teacher's Assignments & Student Work: Opening a Window on Classroom Practice* (CSE 602). Los Ángeles: UCLA.
- Matsumura, L., et al. (2006). *Measuring Reading Comprehension and Mathematics Instruction in Urban Middle Schools: A Pilot Study of the Instructional Quality Assessment* (CSE 681). Los Ángeles: UCLA.
- Mayer, D. (1999). Measuring Instructional Practice: Can Policy Makers Trust Survey Data? *Educational Evaluation and Policy Analysis*, 21(1), 29-45.
- McCaffrey, D.F., et al. (2003). *Evaluating Value-added Models for Teacher Accountability*. Santa Mónica: Rand Co.
- Medley, D.M., y Mitzel, H. (1963). Measuring Classroom Behavior by Systematic Observation. En N.L. Gage (ed.). *Handbook of Research on Teaching* (pp. 247-328). Chicago: Rand McNally.
- Medley, D.M., y Coker, H. (1987). The Accuracy of Principals' Judgments of Teacher Performance. *Journal of Educational Research*, 80, 242-247.
- MET Project (2010a). *A Composite Measure of Teacher Effectiveness*. MET Project Research Paper. Bill & Melinda Gates Foundation.
- MET Project (2010b). *Danielson's Framework for Teaching for Classroom Observations*. Bill & Melinda Gates Foundation.
- MET Project (2010c). *Learning about Teaching: Initial Findings from the Measures of Effective Teaching Project*. MET Project Research Paper. Bill & Melinda Gates Foundation.
- MET Project (2010d). *Overview: Teacher Observation Rubrics*. Measures of Effective Teaching. Teachscape.

- MET Project (2010e). *The MQI Protocol for Classroom Observations*. Bill & Melinda Gates Foundation.
- MET Project (2010f). *The PLATO Protocol for Classroom Observations*. Bill & Melinda Gates Foundation.
- MET Project (2010g). *Validation Engine for Observational Protocols*. Bill & Melinda Gates Foundation.
- MET Project (2010h). *Student Perceptions and the MET Project*. Bill & Melinda Gates Foundation.
- Morduchowicz, A. (2011). *Diseño y experiencias de incentivos salariales docentes* (documento núm. 55). Santiago-Washington: PREAL.
- National Board for Professional Teaching Standards (1989). *Toward high and Rigorous Standards for the Teaching Profession*. Washington: autor.
- NCCTQ. National Comprehensive Center for Teacher Quality (2011). *Guide to Teacher Evaluation Products*. National Comprehensive Center for Teacher Quality. Recuperado el 18 de enero de 2011, de: www.tqsource.org
- OCDE. Organización para la Cooperación y el Desarrollo Económicos (2011). *Establecimiento de un marco para la evaluación e incentivos docentes. Consideraciones para México*. París: autor.
- Odden, A., y Kelley, C. (2002). *Paying Teachers for What they Know and Do: New and Smarter Compensation Strategies to Improve Schools* (2a ed.). Thousand Oaks: Corwin Press.
- Olson, M. (1975). Measurement and Efficiency in Education. En *Educational Indicators: Monitoring the State of Education. Proceedings of the ETS 1975 Invitational Conference* (pp. 41-53). Princeton: Educational Testing Service.
- Pacheco, A. (2009). Mapping the Terrain of Teacher Quality. En D. Gitomer (ed.). *Measurement Issues and Assessment of Teaching Quality* (pp. 160-178). Thousand Oakes: Sage.
- Paige, M. (2012). Using VAM in High-stakes Employment Decisions. *Phi Delta Kappan*, 94(3), 29-32.
- Peterson, K.D. (2000). *Teacher Evaluation. A Comprehensive Guide to New Directions and Practices* (2a ed.). Thousand Oaks: Corwin Press.
- Pianta, R.C., y Hamre, B.K. (2009). Conceptualization, Measurement and Improvement of Classroom Processes: Standardized Observation Can Leverage Capacity. *Educational Researcher*, 38(2), 109-119.
- Porter, A., Youngs, P., y Odden, A. (2001). Advances in Teacher Assessments & Their Uses. En V. Richardson (ed.). *Handbook of Research on Teaching* (4a ed., pp. 295-297). Washington: AERA.
- Potemski, A., et al. (2011). *Alternative Measures of Teacher Performance*. Washington: National Comprehensive Center for Teacher Quality.
- Protsik, J. (1996). History of Teacher Pay and Incentives Reform. *Journal of School Leadership*, 6(3), 265-289.
- Ravela, P., Picaroni, B., y Loureiro, G. (2009, julio). La evaluación de aprendizajes en las aulas de 6° grado en América Latina. *Observatorio Regional de Políticas de Evaluación Educativa* (boletín núm. 12, pp. 1-12). GTEE-PREAL.

- Resnick, L., et al. (2010). The Relationship between Assessment and the Organization and Practice of Teaching. En P. Peterson, E. Baker y B. McGaw (eds.). *International Encyclopedia of Education* (3a ed., vol. 3, pp. 397-402). Nueva York: Elsevier-Academic Press.
- Ritter, G.W., y Shuls, J.V. (2012). If a Tree Falls in a Forest, but no one Hears... *Phi Delta Kappan*, 94(3), 34-38.
- Rivkin, S. (2005). Stricter Regulations or Additional Incentives? The Teacher Quality Policy Dilemma. En C.A. Dwyer. *Measurement and Research in the Accountability Era* (pp. 257-273). Mahwah: Lawrence Erlbaum.
- Robinson, W., y Campbell, J. (2010). Evaluation of Teacher Quality in Practice. En P. Peterson, E. Baker y B. McGaw (eds.). *International Encyclopedia of Education* (3a ed., vol. 3, pp. 674-680). Ámsterdam: Elsevier-Academic Press.
- Rosenshine, B., y Furst, F. (1973). The Use of Direct Observation to Study Teaching. En R.M.W. Travers (ed.). *Second Handbook of Research on Teaching* (pp. 128-183). Chicago: Rand McNally College Publ. Co.
- Rothstein, J. (2011). *Review of "Learning about Teaching: Initial Findings from the MET Project"*. Boulder: National Education Policy Center.
- Rowan, B., Camburn, E., y Correnti, R. (2004). Using Teacher Logs to Measure the Enacted Curriculum in Large-scale Surveys: Insights from the Study of Instructional Improvement. *Elementary School Journal*, 105, 75-102.
- Rowan, B., y Correnti, R. (2009). Studying Reading Instruction with Teacher Logs: Lessons from the Study of Instructional Improvement. *Educational Researcher*, 38(2), 120-131.
- Rubin, D.B., Stuart, E.A., y Zanutto, E.L. (2004). A Potential Outcomes View of Value-added Assessments in Education. *Journal of Educational and Behavioral Statistics*, 29(1), 103-116.
- Sadler, D.R. (1989). Formative Assessment and the Design of Instructional Systems. *Instructional Science*, 18, 119-144.
- Santibáñez, L. (2010). Teacher Incentives. En P. Peterson, E. Baker y B. McGaw (eds.). *International Encyclopedia of Education* (3a ed., vol. 2, pp. 481-488). Oxford-Nueva York: Elsevier-Academic Press.
- Scriven, M. (1967). The Methodology of Evaluation. En R.W. Tyler, R.M. Gagne y M. Scriven (eds.). *Perspectives of Curriculum Evaluation*. Chicago: Rand McNally.
- Sedlak, M. (2008). Competing Visions of Purpose, Practice and Policy: The History of Teacher Certification in the United States. En M. Cochran-Smith et al. (eds.). *Handbook for Teacher Educators: Enduring Questions and Changing Contexts*. Nueva York: Taylor & Francis/Association of Teacher Educators.
- Sharpes, D.K. (1987). Incentive Pay and the Promotion of Teaching Proficiencies. *The Clearing House*, 60, 406-408.
- Shulman, L.S. (1986). Those Who Understand Teach. *Educational Researcher*, 15 (2), 4-14.
- Shulman, L.S. (1987a). Assessment for Teaching: An Initiative for the Profession. *Phi Delta Kappan*, 69, 38-44.
- Shulman, L.S. (1987b). Knowledge and Teaching: Foundations of the New Reform. *Harvard Educational Review*, 57, 1-22.

- Shulman, L.S. (1989). The Paradox of Teacher Assessment. En *ETS, New Directions for Teacher Assessment. Proceedings of the 1988 ETS Invitational Conference* (pp. 13-27). Princeton: autor.
- Shulman, L.S. (2009). Assessment of Teaching or Assessment for Teaching? Reflections on the Invitational Conference. En D. Gitomer (ed.). *Measurement Issues and Assessment for Teaching Quality* (pp. 234-244). Thousand Oakes: Sage.
- Shulman, L.S., y Sykes, G. (1986). *A National Board for Teaching? In Search of Bold Standards*. Carnegie Forum on Education and Economy.
- Shulman, L., Haertel, E., y Bird T. (1988). *Toward Alternative Assessments of Teaching: A Report of Work in Progress*. Stanford Univ. School of Education, Teacher Assessment Project.
- SICI (2013a). *The Standing International Conference of Inspectorates*. Consultado el 7 de febrero de 2013 en la página de la SICI.
- SICI (2013b). *Présentation et missions de l'inspection générale de l'éducation nationale*. Consultado el 7 de febrero de 2013 en la página de la SICI.
- SICI (2013c). *The Inspectorate of Education of Spain*. Consultado el 7 de febrero de 2013 en la página de la SICI.
- SICI (2013d). *The Inspectorate of England*. Consultado el 7 de febrero de 2013 en la página de la SICI.
- Stecher, B., et al., (2006). Using Structured Classroom Vignettes to Measure Instructional Practices in Mathematics. *Educational Evaluation and Policy Analysis*, 28(2), 101-130.
- Stein, M.K., y Matsumura, L.C. (2008). Measuring Instruction for Teacher Learning. En D. Gitomer (ed.). *Measurement Issues and Assessment of Teaching Quality* (pp. 179-205). Thousand Oakes: Sage.
- Stiggins, R.J., y Nancy, F.C., (1992). *In Teacher's Hands: Investigating the Practices of Classroom Assessment*. Albany: State University of New York Press.
- Stigler, J., Gallimore, R., y Hiebert, J. (2000). Using Video Surveys to Compare Classrooms and Teaching Across Cultures: Examples and Lessons from the TIMSS Video Studies. *Educational Psychologist*, 35(2), 87-100.
- Strong, M. (2011). *The Highly Qualified Teacher: What is Teacher Quality and How Do We Measure It?* Nueva York: Teachers College Press.
- Stronge, J.H., Gareis, C., y Little, C. (2006). *Teacher Pay & Teacher Quality: Attracting and Retaining the Best Teachers*. Thousand Oaks: Corwin Press.
- Stufflebeam, D.L., y Nevo, D. (1995). Evaluation of Educational Personnel. En T. Husén y T.S.N. Postlethwaithe (eds.). *The International Encyclopedia of Education* (3a ed., vol. 4, pp. 2123-2132). Oxford/Nueva York: Pergamon.
- Task Force on Teaching as a Profession (1986). *A Nation Prepared: Teachers for the 21st Century*. Carnegie Forum on Education and Economy.
- Vaillant, D. (2012). La gobernanza educativa y los incentivos docentes: Los casos de Chile y Uruguay. *Revista Uruguaya de Ciencia Política*, 21(1), 119-139.
- Van Bruggen, J.C. (2010). *Inspectorates of Education in Europe: Some Comparative Remarks about their Tasks and Work*. Standing International Conference of Inspectorates.

- Van Littlefield, N. (1994). A Yearly Contract with Everybody and his Brother. Durham County, Black Female Public School Teachers, 1885-1927. *Journal of Negro History*, 79, 37-53.
- Watson (1995). School Inspectors and Supervisors. En T. Husén y T.S.N. Postlethwaithe (eds.). *The International Encyclopedia of Education* (3a ed., vol. 9, pp. 5247-5252). Oxford/Nueva York: Pergamon.
- Weiss, E. (2013). *El buen docente y su evaluación* (documento de discusión para el Ciclo Discusiones en torno a la Reforma Educativa). México: Departamento de Investigaciones Educativas.
- Weisberg, D., et al. (2009). *The Widget Effect. Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness* (2a ed.). Washington: The New Teacher Project.
- Wiliam, D. (2007). Content Then Process: Teacher Learning Communities in the Service of Formative Assessment. En Reeves, D. (Ed.). *Ahead of the Curve: The Power of Assessment to Transform Teaching and Learning* (183-204). Bloomington: Solution Tree Press.
- Wilms, W.W., y Chapleu, R.R. (1999). The Illusion of Paying Teachers for Student Achievement. *Education Week*, 19, 34, 48.
- Wilson, S.M. (2009). Measuring Teacher Quality for Professional Entry. En D. Gitomer (ed.). *Measurement Issues and Assessment of Teaching Quality* (pp. 8-29). Thousand Oakes: Sage.

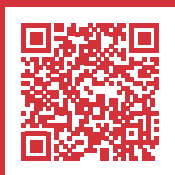


LA EVALUACIÓN DE DOCENTES DE EDUCACIÓN BÁSICA
UNA REVISIÓN DE LA EXPERIENCIA INTERNACIONAL

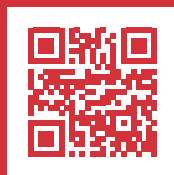
Se terminó de imprimir en junio de 2016 en los talleres
de Impresora y Encuadernadora Progreso, S.A. de C.V. (IEPSA).
En su formación se emplearon las familias tipográficas
Avenir LT Std y The Serif. Esta edición consta
de 1500 ejemplares.



Descargue una
copia digital gratuita



Comuníquese
con nosotros



Visite nuestro
portal