

# DOCUMENTOS

**Nº 46**

**Evaluación docente:  
prácticas vigentes en los  
países de la OCDE y una  
revisión de la literatura**

Marzo 2010

**Marlène Isoré**

Este informe fue preparado por **Marlène Isoré**, estudiante de postgrado del Institut d'Études Politiques de Paris (Sciences Po), Francia, durante una pasantía en la División de Políticas de Educación y Capacitación, Dirección de Educación, OCDE, durante el período junio-septiembre de 2008. PREAL agradece a la OCDE la autorización para traducirlo y publicarlo en español.

Las opiniones vertidas en este documento no representan necesariamente al PREAL ni a las instituciones que lo patrocinan.

Publicado originalmente por OECD en inglés como:  
Isoré, M. (2009), “Teacher Evaluation: Current Practices in OECD Countries and a Literature Review”,  
OECD Education Working Paper No. 23  
© 2009 OECD

© 2010, PREAL, Chile, para esta edición en español.  
Publicado en acuerdo con OECD, París.  
La calidad de la traducción al español y su coherencia con el texto original es de responsabilidad de PREAL, Chile.  
Traducción: Maritza Blajtrach

Este documento puede ser descargado de: [www.preal.org/publicacion.asp](http://www.preal.org/publicacion.asp)  
Si desea solicitar copias adicionales, contacte a PREAL  
a través de CINDE o del Diálogo Interamericano:

CINDE  
Santa Magdalena 75, piso 10, oficina 1002, Providencia  
Santiago, Chile  
Tel.: 56-2-3344302

INTER-AMERICAN DIALOGUE  
1211 Connecticut Ave., NW, Suite 510  
Washington, D.C., 20036  
Tel.: 202-822-9002

E-mail: [infopreal@preal.org](mailto:infopreal@preal.org)

ISSN 0718-6002  
Primera edición  
Publicado en Chile

Diseño e Impresión:  
Editorial San Marino  
[esm.cl](http://esm.cl)

# ÍNDICE

<b>1. Introducción</b> .....	4
<b>2. Principales características de los esquemas de evaluación del desempeño docente</b> .....	5
2.1 Propósitos de la evaluación .....	5
2.1.1 La evaluación sumativa y la garantía de la calidad .....	5
2.1.2 Evaluación formativa y desarrollo profesional .....	6
2.2 Principales elementos de los esquemas de evaluación docente .....	7
2.2.1 Actores involucrados en la concepción e implementación de los sistemas de evaluación .....	7
2.2.2 Alcance de la evaluación y maestros evaluados .....	9
2.2.3 Criterios y estándares .....	10
2.2.4 Instrumentos de recopilación de datos .....	12
2.2.5 Evaluadores .....	13
2.3 Vínculos con el reconocimiento y las retribuciones .....	15
2.4 Vínculos con las oportunidades de desarrollo profesional y las prioridades más generales de las escuelas .....	16
<b>3. Ventajas y desventajas de los diferentes enfoques con respecto a la evaluación docente</b> .....	17
3.1 Ventajas y desventajas del uso de los indicadores de rendimiento escolar como medida del desempeño docente .....	17
3.2 Diseño de un conjunto coherente de métodos e instrumentos alineado con el propósito enfatizado de evaluación docente .....	19
3.3 Dificultades en la implementación de los esquemas de evaluación del desempeño docente .....	21
3.3.1 Conflictos de interés entre los diferentes actores .....	21
3.3.2 Cómo superar los obstáculos .....	23
<b>4. Evidencia empírica de los efectos de los esquemas de evaluación del desempeño docente</b> .....	24
4.1 Evidencia cuantitativa .....	24
4.1.1 Dificultades para medir la calidad del desempeño docente y el impacto de la evaluación del desempeño docente .....	24
4.1.2 Evidencia empírica mixta con respecto a los sistemas de evaluación del desempeño docente .....	25
4.2 Evidencia cualitativa .....	27
<b>5. Conclusión</b> .....	29
<b>Anexo 1: Marco conceptual para la evaluación del desempeño docente</b> .....	30
<b>Anexo 2: Ejemplos de sistemas de evaluación del desempeño docente en los países de la OCDE</b> .....	31
1. Evaluación del desempeño docente con propósitos sumativos y vínculos con la remuneración: .....	31
2. Evaluación docente con propósitos formativos y como parte de las políticas escolares más generales .....	31
2a. Finlandia [UNESCO, 2007] .....	31
2b. Inglaterra [Ofsted, 2006; TDA, 2007] .....	32
3. Conciliación de los propósitos sumativos y formativos en un enfoque integral: Chile [Ávalos y Assael, 2006] .....	32
4. Evaluación docente derivada de procedimientos burocráticos: Francia [Haut Conseil de l'évaluation de l'école, 2003; Pochard, 2008] .....	33
<b>Referencias</b> .....	34

## RESUMEN

En este informe se analizan los principales problemas relativos a la evaluación del desempeño docente en la educación primaria y secundaria mediante una revisión de la literatura reciente y un examen de las prácticas vigentes en los países de la OCDE. En primer lugar, se presenta un marco conceptual que destaca las principales características de los esquemas de evaluación docente. En particular, se recalca la importancia de aclarar los propósitos de la evaluación, ya sean sumativos, cuando el objetivo es garantizar que se están aplicando prácticas que favorecen el aprendizaje de los estudiantes, o formativos, cuando el objetivo es la formación profesional continua. Además, se consideran los diversos criterios e instrumentos utilizados comúnmente para evaluar a los maestros, como también los actores generalmente involucrados en el proceso y las posibles consecuencias para la vida profesional de los docentes. En segundo lugar, se aborda una serie de puntos conflictivos, incluyendo la conveniencia de utilizar los indicadores de rendimiento escolar para medir el desempeño de los docentes, las ventajas y desventajas de diferentes enfoques dependiendo del propósito enfatizado y las restricciones de recursos, las dificultades de implementación resultantes de los diferentes intereses de los actores y las posibles maneras de superar estos obstáculos. Finalmente, se examina la evidencia empírica actual, que apunta a resultados mixtos a consecuencia de las dificultades para medir los efectos de los esquemas de evaluación existentes sobre la calidad de la enseñanza, la motivación de los maestros y el aprendizaje de los estudiantes. Como conclusión, se consideran las circunstancias bajo las cuales los sistemas de evaluación del desempeño docente parecen ser más eficaces, equitativos y confiables. El desarrollo de un enfoque integral para evaluar a los maestros es crítico para conciliar las demandas de mejores prácticas educacio-

nales con la apropiación del proceso por parte de los maestros, así como también para aumentar el decisivo atractivo y reconocimiento de la profesión docente.

## 1. INTRODUCCIÓN

En este informe se examina la literatura académica y de políticas actual sobre evaluación del desempeño docente en la educación primaria y secundaria. Se actualiza y expande la sección correspondiente incluida en “Los docentes son importantes: cómo atraer, formar y retener a los docentes eficientes” (*Teachers Matter: Attracting, Developing and Retaining Effective Teachers*), publicado por la OCDE en 2005.

La evaluación de la práctica y el desempeño de los docentes no es una preocupación reciente. Como resultado de las demandas por una mejor calidad de la educación, muchos países han establecido alguna forma de evaluación del desempeño docente. Sin embargo, la evaluación docente ha sido siempre un tema muy controvertido, con evidencia empírica mixta con respecto a sus efectos en el aprendizaje de los estudiantes y con conflictos de interés entre los principales actores de los sistemas educacionales. Por consiguiente, la evaluación ha sido con frecuencia un ejercicio sinsentido, derivado de los rituales burocráticos requeridos en las escuelas y sobrellevado tanto por los docentes como por los evaluadores (Danielson, 2001; Holland, 2005; Marshall, 2005). Solo recientemente algunos países han mostrado un creciente interés por establecer sistemas de evaluación como un elemento integral de las políticas docentes y escolares más generales (Peterson, 2006; TDA 2007a).

Los esquemas de evaluación docente vigentes en los sistemas educacionales de la OCDE

adoptan múltiples formas. El alcance y los métodos de la evaluación docente, los criterios y estándares aplicados y los instrumentos de recopilación de datos difieren ampliamente de un país a otro, dependiendo del contexto educacional y la tradición, de los actores involucrados en el diseño e implementación del sistema de evaluación y del propósito de evaluación enfatizado. Las consecuencias de los procesos de evaluación para las carreras profesionales de los docentes también son diversas. Si bien el escalón único y la estructura salarial única siguen aplicándose en forma generalizada, diversos países han intentado vincular su sistema de evaluación docente ya sea al reconocimiento y las retribuciones, financieras o de otro tipo, o a las oportunidades de desarrollo profesional.

Este informe consta de tres secciones más. En la sección 2 se examinan las principales dimensiones de los esquemas de evaluación docente encontradas en la literatura o en los sistemas educacionales de la OCDE. En la sección 3 se analizan las ventajas y desventajas de los diferentes enfoques, como también las dificultades planteadas por la implementación de programas de evaluación eficaces debido a los diferentes puntos de vista de los actores. Finalmente, en la sección 4 se resume la evidencia existente con respecto a los efectos de los sistemas de evaluación docente.

## **2. PRINCIPALES CARACTERÍSTICAS DE LOS ESQUEMAS DE EVALUACIÓN DEL DESEMPEÑO DOCENTE**

En esta sección se describen las principales dimensiones de los esquemas de evaluación del desempeño docente encontradas en la literatura y en los sistemas educacionales de la región de la OCDE.

### **2.1 PROPÓSITOS DE LA EVALUACIÓN**

La evaluación del desempeño docente tiene dos propósitos principales. Por una parte, garantizar que los maestros se desempeñen de la mejor manera posible para favorecer el aprendizaje de los estudiantes. Por otra, mejorar la propia práctica de los maestros a través de la identificación de sus fortalezas y deficiencias con miras a la formación profesional continua. Estos dos enfoques están referidos a dos formas de evaluación de naturaleza diferente, sumativa y formativa, respectivamente.

#### **2.1.1 La evaluación sumativa y la garantía de la calidad**

Si el objetivo final de los sistemas educacionales es favorecer el aprendizaje de todos los estudiantes y si el desempeño y la práctica de los docentes constituyen el principal factor en este propósito, la evaluación docente podría considerarse como un mecanismo de garantía de la calidad (Danielson y McGreal, 2000; Kleinhenz e Ingvarson, 2004). Suponiendo que la calidad de los docentes y la calidad de la educación son importantes, un proceso de evaluación idealmente debería estar orientado tanto a la eficiencia educacional –garantizando que la enseñanza cumpla con los estándares académicos requeridos para que los estudiantes vivan en sociedades del conocimiento– como a la equidad educacional, garantizando que todos los estudiantes tengan acceso a las oportunidades de progreso escolar independientemente de sus antecedentes. Por lo tanto, la evaluación *sumativa* del desempeño docente es una forma de evaluar que los docentes estén adoptando las acciones y prácticas ‘óptimas’ para mejorar los resultados de los estudiantes.

El desarrollo de una evaluación *sumativa* es la manera más visible y reconocible de evaluar a una persona y consiste en realizar informes resumidos de las capacidades de un maestro a través de exámenes, con el fin de medir sus aptitudes y conocimientos, garantizar el cumplimiento de los estándares requeridos o certificar el nivel de desempeño con miras a un reconocimiento inmediato. La evaluación docente sumativa entrega información esencial con respecto a las prácticas que está aplicando y al desempeño que está teniendo el maestro que está siendo evaluado en relación con los estándares de ‘buena’ enseñanza establecidos. En consecuencia, la evaluación sumativa es una fuente de documentación indispensable para responsabilizar a los maestros por su desempeño profesional. Por ejemplo, Stronge y Tucker (2003) recalcan la necesidad de un mecanismo de garantía de calidad como este: “El propósito de responsabilidad por los resultados refleja un compromiso con los importantes objetivos profesionales de competencia y desempeño de calidad. Esta función de responsabilidad por los resultados (...) está relacionada con la evaluación de la eficiencia de los servicios educacionales”.

La necesidad de mecanismos de responsabilidad por los resultados en la docencia proviene de la existencia de relaciones asimétricas, típi-

cas del 'problema de la relación mandante-mandatario' tan conocido para las teorías de las ciencias económicas y políticas. El 'mandante' carece de información crucial para saber si su empleado, el 'mandatario', se está desempeñando de acuerdo a sus expectativas de resultados. En nuestro caso particular, los padres, las autoridades a cargo de la calidad de la educación e incluso los directores de los establecimientos escolares solo tienen medios limitados para conocer el grado de concordancia entre el desempeño de los maestros y sus expectativas de aprendizaje de los estudiantes. En palabras de Mizala y Romaguera (2004): "La responsabilidad por los resultados es fundamental, debido a que existe una brecha de información que separa a las escuelas de las familias. Es difícil para las familias obtener información actualizada y pertinente con respecto a lo que está ocurriendo en sus escuelas y las escuelas no reciben necesariamente incentivos para entregar información a los padres. Además, la capacidad de las familias para obtener información con respecto a las escuelas varía, dependiendo de su nivel cultural y socioeconómico". Considerando esta información asimétrica, Casson (2007) sostiene que, en ausencia de mecanismos de incentivo, los maestros tienen el incentivo de realizar menos esfuerzo (dado el costo asociado a una mayor cantidad de trabajo), debido a que "el distrito escolar no puede distinguir entre el bajo rendimiento de los estudiantes producto de una falta de esfuerzo del maestro y el bajo rendimiento de los estudiantes producto de la escasa habilidad de los estudiantes". Si bien Casson argumenta a favor de medidas tendientes a armonizar el desempeño de los maestros con los intereses de las autoridades *a posteriori* para superar este problema, un sistema de evaluación sumativa bien diseñado podría ser suficiente como para motivar a los maestros a adoptar las mejores prácticas, debido a que esto permite cerrar la brecha de información *a priori*.

Además del propósito informativo en sí mismo, los resultados de las evaluaciones sumativas permiten la toma de decisiones con consecuencias con respecto al maestro que está siendo evaluado. De acuerdo a Ávalos y Assael (2006): "La mayor parte de las formas de evaluación se justifican debido a que se requiere información de diagnóstico o debido a que entregan evidencia para la toma de decisiones. Esto también es válido para la evaluación del desempeño docente". La evaluación de los maestros en relación con criterios específicos per-

mite realizar comparaciones y estas últimas son útiles para las decisiones con respecto a la contratación y permanencia en los cargos, las oportunidades de ascenso o, en condiciones particulares, las medidas a aplicar en el caso de los maestros ineficientes.

La evaluación sumativa del desempeño docente también puede utilizarse como base para el reconocimiento y la retribución del trabajo de los maestros. Existen inquietudes con respecto a la imagen y el prestigio de la profesión docente en varios países de la OCDE, incluyendo el sentimiento de los maestros de que su trabajo está subvalorado. La evaluación ofrece oportunidades para reconocer y retribuir la competencia y el desempeño de los docentes, aspectos esenciales para retener a los docentes eficientes en las escuelas, como también para hacer de la docencia una elección atractiva como carrera profesional (OCDE, 2005). Por ejemplo, la Junta Nacional Norteamericana para los Estándares Profesionales Docentes (*US National Board for Professional Teaching Standards* - NBPTS) ha desarrollado "rigurosos estándares profesionales que reflejan lo que los buenos maestros deben saber y ser capaces de hacer" como base para un sistema nacional voluntario de certificación de los docentes que cumplen con estos estándares. Desde su creación en 1987, más de 64.000 docentes norteamericanos han sido reconocidos por su desempeño sobresaliente.

En su forma sumativa, la evaluación responde en primer lugar a las necesidades de garantizar que la enseñanza se oriente al rendimiento académico de los estudiantes. También brinda oportunidades de reconocimiento social de las habilidades de los maestros y su compromiso con su trabajo. Estas son dos inquietudes principales en nuestras sociedades del conocimiento.

### **2.1.2 Evaluación formativa y desarrollo profesional**

Además de los propósitos de responsabilidad por los resultados y reconocimiento, la evaluación docente puede aplicarse con el fin de mejorar la práctica de un maestro. La evaluación *formativa* se refiere a una evaluación cualitativa de la práctica actual de un maestro, orientada a identificar sus fortalezas y deficiencias y a entregar oportunidades de desarrollo profesional adecuadas en las áreas en las cuales se requiere perfeccionamiento. Como explican Stronge y

Tucker (2003): “El propósito de mejoramiento del desempeño está relacionado con la dimensión de desarrollo personal e implica ayudar a los maestros a aprender y a reflexionar acerca de su práctica y a mejorarla. Esta función de mejoramiento generalmente es considerada de naturaleza formativa y sugiere la necesidad de perfeccionamiento y formación profesional continua”. Al contrario de una evaluación sumativa diseñada para formular juicios acerca de un desempeño (evaluación *de* la enseñanza), la función de una evaluación formativa es determinar las maneras de mejorar la práctica actual (evaluación *orientada a* la enseñanza).

La evaluación formativa es un proceso a través del cual los evaluadores entregan retroalimentación constructiva al maestro, indicando el nivel al cual se está desempeñando en cada uno de los criterios pertinentes y sugiriendo maneras de mejorar su práctica. Las conversaciones con los evaluadores o con colegas comprometen a los maestros en una auto-reflexión acerca de su trabajo. En palabras de Danielson y McGreal (2000): “A medida que los maestros consideran la articulación de diferentes componentes de la docencia y sus elementos y comparan sus impresiones y prácticas con las de los demás, intercambian técnicas y aprenden nuevas estrategias de sus colegas. Estas conversaciones son enriquecedoras; están centradas en la calidad de la enseñanza y aportan mucho al aprendizaje profesional de los participantes”. El empoderamiento de cada uno de los docentes en el perfeccionamiento de sus propias habilidades trasciende el propósito de garantía de la calidad inherente a la evaluación docente.

Además, los resultados de la evaluación formativa permiten que las escuelas adapten sus programas de desarrollo profesional a las necesidades de sus docentes de acuerdo a sus objetivos educacionales. Las escuelas pueden aprender a partir de las fortalezas de los maestros eficientes –destacadas por las evaluaciones formativas– e implementar programas de desarrollo profesional que respondan a sus deficiencias. En Finlandia, el director de un establecimiento escolar es el líder pedagógico responsable de los maestros que trabajan en su establecimiento y de la implementación de las medidas necesarias para mejorar la calidad de la enseñanza. En consecuencia, la mayoría de las escuelas finlandesas tienen un sistema que incluye debates anuales destinados a evaluar el cumplimiento

por parte de los maestros de los objetivos individuales establecidos durante el año anterior y analizar los objetivos y necesidades individuales para el año siguiente (UNESCO, 2007).

Asimismo, las instituciones a cargo de la formación docente inicial también pueden beneficiarse a partir de la retroalimentación provista por las evaluaciones formativas. Pecheone y Chung (2006) describen la Evaluación del Desempeño para los Maestros de California (*Performance Assessment for California Teachers - PACT*), un sistema de evaluación de las competencias de los estudiantes de pedagogía para efectos de la obtención de la licenciatura, como “una poderosa herramienta para el aprendizaje de los maestros y el mejoramiento de los programas”. En efecto, agregan que este sistema ha generado muchos debates profesionales “acerca de las características de una enseñanza eficaz y acerca de lo que los egresados deberían saber y ser capaces de hacer”, lo que, a su vez, ha llevado a los programas a “reexaminar la manera en que apoyan y preparan a los postulantes a la docencia”.

En su modalidad formativa, la evaluación puede considerarse como una base para el mejoramiento de la docencia y las oportunidades de desarrollo profesional durante toda la vida. Los aspectos sumativos y formativos de la evaluación del desempeño docente suelen ser propósitos conflictivos, pero no necesariamente incompatibles. En la práctica, los países rara vez utilizan una forma pura de modelo de evaluación docente, sino que más bien aplican una combinación singular que integra propósitos y metodologías múltiples (Stronge y Tucker, 2003).

## **2.2 PRINCIPALES ELEMENTOS DE LOS ESQUEMAS DE EVALUACIÓN DOCENTE**

En esta sección se resumen los elementos involucrados en los sistemas de evaluación docente, tales como los actores que participan en el diseño e implementación del proceso, el alcance de la evaluación, los instrumentos y métodos de recopilación de datos y los criterios y estándares aplicados para evaluar a los maestros.

### **2.2.1 Actores involucrados en la concepción e implementación de los sistemas de evaluación**

**Gobiernos.** Los gobiernos desempeñan una función principal en la concepción de los esquemas

de evaluación, dado que establecen los objetivos en términos de indicadores de aprendizaje a nivel nacional, frecuentemente a través de la legislación. Por ejemplo, en Estados Unidos, la Ley 'Ningún Niño Debe Quedar Atrás' (*No Child Left Behind* - NCLB), promulgada en 2001, pretende mejorar el desempeño de las escuelas primarias y secundarias norteamericanas elevando los estándares de responsabilidad por los resultados para los estados, los distritos escolares y las escuelas. Los sistemas de evaluación eficientes deben estar orientados al logro de los objetivos nacionales. Además, algunas veces los gobiernos desempeñan un rol directo en la implementación y supervisión de los procedimientos de evaluación docente. La extensión de esta función depende del grado de descentralización del país (UNESCO, 2007). En este sentido, Francia es un caso paradigmático: la mayor parte de las decisiones legales se toman a nivel central, incluyendo las decisiones relativas a las escuelas públicas; por lo tanto, el Ministerio de Educación es el encargado de determinar los diferentes aspectos del sistema de evaluación. A nivel más general, las autoridades nacionales son actores de mayor importancia en los países en los cuales los docentes son funcionarios públicos.

**Autoridades locales.** Las autoridades locales encargadas de las políticas educacionales generalmente tienen uno de los dos deberes siguientes en relación con la evaluación docente. En algunos países de la OCDE, las autoridades locales son responsables del logro de los objetivos nacionales y, por lo tanto, implementan procedimientos considerados deseables para garantizar la calidad educacional de los establecimientos escolares bajo su responsabilidad. Por ejemplo, Heneman *et al.* (2006) resumen las características de los modelos de evaluación docente de cuatro distritos norteamericanos (Cincinnati, Washoe, Coventry y Vaughn). En otros países, las autoridades locales pueden tomar parte en la concepción del esquema de evaluación, pero están, sobre todo, a cargo de la implementación y supervisión de las medidas de evaluación docente decididas por el Estado central. Por ejemplo, las municipalidades chilenas brindaron asesoría en el diseño del proceso de evaluación, pero actualmente todas aplican el mismo sistema nacional, que fue promulgado por ley en 2004 (Ávalos y Assael, 2006).

**Directivos escolares.** Mientras más descentralizado es un país, mayor es la importancia del rol que cumplen los directivos escolares en el

diseño e implementación del proceso de evaluación. En Finlandia, donde el sistema educacional se caracteriza por un muy alto grado de autonomía escolar, todas las decisiones relativas a los docentes (incluida la evaluación) se toman en las escuelas mismas (UNESCO, 2007). El rol de los directivos escolares como evaluadores se analizará más adelante.

**Investigadores educacionales y maestros experimentados.** Para el diseño del sistema de evaluación se puede solicitar la asesoría de investigadores y maestros en su calidad de expertos en la materia. Estos profesionales se encuentran en una buena posición como para saber cuáles son las 'buenas' prácticas docentes, ya sea como resultado de sus estudios o su propia experiencia como docentes, y pueden ayudar a identificar los criterios e instrumentos pertinentes para evaluar a los maestros (Ingvarson, Kleinhenz y Wilkinson, 2007). Los integrantes de la NBPTS, constituida principalmente por maestros, "examinaron los pros y los contras de diferentes métodos de investigación y luego aplicaron sus propias experiencias a lo que oyeron y aprendieron, siempre teniendo en consideración la intersección de los datos empíricos a gran escala, su propia formación y desarrollo profesional como maestros expertos y la naturaleza de los estudiantes a los que enseñan y sirven. Deliberaron y debatieron entre ellos y luego consultaron a otros colegas con el fin de generar nuevas perspectivas e ideas" (NBPTS, 2007).

**Sindicatos docentes.** Los sindicatos docentes también pueden ser consultados para el diseño e implementación de procedimientos relativos a las prácticas y dificultades cotidianas de los maestros. Se supone que los sindicatos docentes representan todos los intereses de los maestros, cualquiera sea su nivel de desempeño. Suponiendo que los maestros podrían mostrarse renuentes o temerosos frente a algunos aspectos particulares de los esquemas de evaluación docente, los sindicatos docentes son indispensables para diseñar un proceso que considere sus intereses y se traduzca en un amplio consenso. Heneman *et al.* (2006) sostienen que deben incorporarse mecanismos para disminuir la resistencia en el diseño inicial del plan. Estos incluyen la comunicación amplia y continua con los docentes y los administradores. Agregan que se requiere un compromiso con respecto a una transformación de la manera en que se define,

mide y apoya el desempeño docente, y que dicho compromiso debe considerar las aprensiones de los docentes y de los administradores.

**Los padres.** Es poco frecuente que los padres participen directamente en el diseño o implementación de los sistemas de evaluación docente, dado que sus intereses educacionales están representados por las autoridades nacionales y locales mencionadas anteriormente. Su rol como evaluadores se analizará más adelante.

### **2.2.2 Alcance de la evaluación y maestros evaluados**

Los procedimientos de evaluación docente no son necesariamente aplicables a todos los maestros de un país; por el contrario, el alcance de la evaluación y los maestros sujetos a la evaluación difieren considerablemente en los distintos sistemas educacionales de la OCDE. Las principales diferencias son las siguientes.

**Procedimientos regionales.** Los mismos procedimientos no son necesariamente aplicables a todo el país, sino que pueden variar de acuerdo a la región considerada. Lo más probable es que los procedimientos varíen a nivel regional cuando la estructura federal o el alto grado de descentralización del país lo permiten. Por ejemplo, en Alemania, los ministerios a cargo de la educación en cada *Land* (estado) determinan sus propias orientaciones para todos los aspectos de la evaluación docente: las personas a cargo del proceso, los criterios de evaluación, el tiempo destinado a la evaluación, los instrumentos de recopilación de datos y las consecuencias de los resultados de la evaluación (UNESCO, 2007).

**Tipo de establecimiento escolar.** El sistema podría estar limitado a las escuelas públicas, pero también podría ser aplicable a algunas escuelas privadas, particularmente a establecimientos al menos parcialmente subsidiados por el Estado, si bien de propiedad y administración privada. Por ejemplo, la *Política de Perfeccionamiento, Supervisión y Evaluación Docente* de la provincia canadiense de Alberta es igualmente aplicable a las escuelas *charter* (escuelas con una organización semiautónoma, pero con financiamiento completamente público) y a las escuelas privadas acreditadas (solo subsidiadas parcialmente con fondos públicos) (Alberta Education, 1996, 2003).

**Nivel de experiencia del maestro.** Las evaluaciones docentes pueden diferir dependiendo del nivel de experiencia del maestro. Por ejemplo, en Inglaterra, las características de los maestros se definen en cada etapa de la carrera profesional (TDA, 2007b), dado que sus funciones y responsabilidades evolucionan durante toda la carrera. Sin embargo, muchos países de la OCDE no diferencian a los maestros de acuerdo a su nivel de experiencia una vez que han obtenido la permanencia en el cargo: “La docencia, a diferencia de todas las demás profesiones, plantea las mismas exigencias a los novatos que a los profesionales experimentados. Al momento en que los maestros de primer año ingresan a sus primeras aulas, se les exige el mismo nivel –y están sujetos a los mismos procedimientos– que sus colegas más experimentados. (...) A pesar de que el distrito escolar debe garantizar que todos los maestros (incluyendo los principiantes) tengan al menos un cierto nivel de competencia, podrían existir algunas variaciones en los procedimientos aplicados a los novatos y a sus colegas más experimentados” (Danielson y McGreal, 2000). No existe consenso con respecto a la manera de diferenciar la evaluación de acuerdo al nivel de experiencia. Por ejemplo, puede argumentarse que los estudiantes de pedagogía deben ser sometidos a evaluaciones más frecuentes y más completas, puesto que deben demostrar la práctica y el desempeño adecuados para obtener su licenciatura. Darling-Hammond *et al.* (2004) señalaron que los esfuerzos de la NBPTS por identificar y reconocer la buena enseñanza motivaron la creación de un sistema similar para evaluar a los estudiantes de pedagogía en California (sistema PACT). Por otra parte, estas evaluaciones podrían no ser necesarias, considerando que el programa de pasantías e inducción para los maestros principiantes les ofrece oportunidades formativas. En consecuencia, la Federación Norteamericana de Maestros recalca que más de la mitad de los estados de Estados Unidos no exigen que el docente haya cumplido un año o dos de enseñanza exitosa para obtener su licenciatura (AFT, 2001). Algunos sostienen que las evaluaciones sumativas deberían ser relativamente más importantes para los maestros experimentados, dado que se supone que comprenden mejor lo que se espera de un ‘buen’ maestro (Danielson y McGreal, 2000), mientras que otros recalcan que su primera necesidad es la retroalimentación formativa para mantener intacta su motivación (Day y Gu, 2007).

**Periodicidad de la evaluación.** Un aspecto principal a considerar es si las evaluaciones formales son parte del trabajo regular del maestro o deben realizarse en ocasiones especiales (y lo contrario es válido para las evaluaciones informales). En Holanda, el 38% de las escuelas primarias y el 62% de las secundarias evalúan a sus maestros de manera regular; en la mayoría de los casos, los maestros son evaluados anualmente (UNESCO, 2007). En otros países, existe un proceso obligatorio de evaluación formal solo cuando los maestros han sido objeto de un reclamo. En Italia, una vez que se otorga la permanencia en el cargo, pueden realizarse evaluaciones formales cuando los administradores de la escuela inician un procedimiento de ‘exoneración de los servicios’, debido a que se ha observado una práctica inadecuada o un desempeño insuficiente durante un período significativo; un maestro también puede solicitar ser evaluado para efectos de rehabilitación si se encuentra sujeto a una sanción disciplinaria (UNESCO, 2007).

**Evaluaciones obligatorias versus evaluaciones voluntarias.** Además de los procedimientos de evaluación obligatorios, algunos países ofrecen a los maestros la posibilidad de ser evaluados en forma voluntaria con el fin de postular a un aumento de su remuneración o a un cargo de más alto nivel. En España, por ejemplo, un maestro de enseñanza secundaria puede ser ascendido excepcionalmente después de una revisión de los resultados de su evaluación por un consejo profesional, el cuerpo de catedráticos (UNESCO, 2007). Es posible combinar las evaluaciones obligatorias formales con evaluaciones voluntarias e informales, puesto que su propósito suele no ser idéntico. En efecto, un procedimiento de evaluación formal está más probablemente destinado a propósitos sumativos mínimos, mientras que un procedimiento de evaluación informal ofrece las mayores oportunidades formativas y un procedimiento voluntario es preferido para propósitos sumativos consecuenciales. Por ejemplo, en Chile coexisten dos formas de evaluación: una evaluación periódica obligatoria para todos los maestros de las escuelas públicas y una evaluación voluntaria para efectos de ascenso (Ávalos y Assael, 2006).

**Implementación piloto.** El alcance de la evaluación puede limitarse a una implementación piloto durante algunos años antes de la plena implementación. El sistema PACT de California

se inició con dos años de implementación piloto (en 2002-2003 y 2003-2004), durante los cuales se evaluaron áreas pedagógicas restringidas con el fin de validar los métodos utilizados y mejorar el programa para los años siguientes (Pecheone y Chung, 2006). Cuanto más se vean afectadas las vidas profesionales de los maestros por el proceso de evaluación, más importante será la implementación piloto.

### 2.2.3 Criterios y estándares

Un esquema de evaluación docente equitativo y confiable requiere criterios y estándares para evaluar a los maestros en relación con lo que se considera una ‘buena’ enseñanza. Deben determinarse las competencias y responsabilidades de los docentes con el fin de concebir una definición integral de lo que los maestros deberían saber y ser capaces de hacer en el ejercicio de su profesión. Un aporte de referencia en esta área es el *Marco para la Enseñanza* de Danielson (1996, 2007), articulado para servir al mismo tiempo como “un *mapa de ruta* para guiar a los maestros principiantes a través de sus experiencias iniciales en las aulas, una estructura para ayudar a que los profesionales experimentados se vuelvan más eficaces y un medio para focalizar las iniciativas de mejoramiento”.

El *Marco* agrupa las responsabilidades de los maestros en cuatro grandes dominios, que luego se dividen en componentes:

- **Planificación y preparación:** demostrar conocimientos de los contenidos y de pedagogía, demostrar conocimiento de los estudiantes, establecer los objetivos pedagógicos, diseñar una instrucción coherente, evaluar el aprendizaje de los estudiantes.
- **El ambiente del aula:** crear un ambiente de respeto e interacción, establecer una cultura propicia para el aprendizaje, manejar los procedimientos de aula, manejar la conducta de los estudiantes y organizar el espacio físico.
- **Instrucción:** comunicarse con los estudiantes en forma clara y precisa, utilizar técnicas de interrogación y debate, comprometer a los estudiantes en el aprendizaje, entregar retroalimentación a los estudiantes, demostrar flexibilidad y sensibilidad.
- **Responsabilidades profesionales:** reflexionar sobre la enseñanza, mantener re-

gistros rigurosos, comunicarse con las familias, participar activamente en el establecimiento escolar y el distrito escolar, realizar actividades de perfeccionamiento y desarrollo profesional, mostrar profesionalismo.

Cada uno de estos componentes consiste en diversos elementos que deben ser evaluados. Por ejemplo, el conocimiento de los estudiantes por parte del maestro comprende elementos tales como el conocimiento de las características de los grupos etarios, el conocimiento de los diversos estilos de aprendizaje de los estudiantes, etc. Cada elemento de un componente está asociado a cuatro niveles de desempeño: 'insatisfactorio', 'básico', 'competente' y 'destacado'.

Danielson sostiene que los niveles de desempeño son especialmente útiles en la supervisión y evaluación, pero también pueden emplearse para ayudar a la autoevaluación o apoyar las relaciones de tutoría o *coaching*, para informar un debate profesional y para sugerir las áreas que requieren perfeccionamiento. En consecuencia, el *Marco* puede servir tanto para efectos sumativos como formativos. Danielson también advierte contra los posibles usos inadecuados de los componentes, señalando que, aunque los componentes sean genéricos y hayan sido diseñados para ser aplicables a cualquier situación pedagógica, sus manifestaciones reales diferirán en los diversos contextos. Por ende, los evaluadores deben examinar la aplicabilidad y ponderación de cada componente, junto con trasladar los elementos específicos observables a contextos particulares.

Kleinhenz e Invargson (2004) advierten de "la ausencia de estándares que describan adecuadamente el trabajo de la docencia – qué es lo que se puede esperar que los docentes sepan y sean capaces de hacer en dominios de práctica específicos", lo que necesariamente se traduce en un deficiente "núcleo técnico de conocimientos y habilidades docentes". Agregan que "actualmente existe consenso con respecto al hecho de que estándares integrales, congruentes y específicos a cada dominio constituyen la única base confiable para la formulación de juicios útiles con respecto a la competencia docente". Entre los conjuntos de estándares que se utilizan actualmente para evaluar a los maestros, algunos fueron propuestos antes del *Marco* de Danielson, pero están referidos a trabajadores en situaciones más específicas. Por ejemplo, los

estándares de la NBPTS están referidos a maestros competentes que se someten voluntariamente a una evaluación para ser reconocidos por el alto nivel de desempeño que han adquirido a través de su experiencia docente. Por el contrario, en el caso de los maestros principiantes, el Nuevo Consorcio Interestatal de Evaluación y Apoyo Docente (*Interstate New Teacher Assessment and Support Consortium* - INTASC) desarrolló estándares de conocimientos y habilidades docentes para los sistemas de titulación de los maestros (CCSSO, 1992).

Una serie de sistemas de evaluación docente de Estados Unidos han establecido una lista de criterios basados en el *Marco para la Enseñanza* de Danielson. Cuatro distritos norteamericanos –Cincinnati, Washoe, Coventry y Vaughn– adoptaron versiones adaptadas del modelo de competencias del *Marco* (Milanowski, 2004; Borman y Kimball, 2005; Heneman *et al.*, 2006). Lo mismo hizo la provincia de Quebec en Canadá. Los cuatro dominios y veinte criterios de evaluación establecidos en Chile se inspiraron en gran medida en el *Marco* (Ávalos y Assael, 2006). El análisis de la UNESCO de los sistemas de evaluación docente europeos y latinoamericanos da énfasis al conocimiento de los contenidos, las habilidades pedagógicas, las habilidades para evaluar a los estudiantes y las responsabilidades profesionales ante el establecimiento escolar, los estudiantes y sus familias como dominios clave para evaluar a los docentes. Cabe señalar que el análisis no menciona la participación en programas de desarrollo profesional como un estándar docente común en los sistemas europeos, con el consiguiente riesgo de subestimar el compromiso y disposición de los maestros con respecto al mejoramiento de su propia práctica. Sin embargo, recientemente Inglaterra implementó un marco para los estándares profesionales similar al de Danielson, que incluye criterios de desarrollo profesional para los cinco niveles de desempeño docente (la certificación como Maestro Calificado, los maestros incluidos en el escalafón principal, los Maestros Post-Umbrales, los Maestros Excelentes y los Maestros con Habilidades Avanzadas) (TDA, 2007b).

Entre los ocho estándares (que incluyen 42 criterios) del Estado de Iowa, los seis siguientes son muy afines al *Marco* de Danielson: competencia en el conocimiento de los contenidos, competencia en planificación y preparación para la instrucción, métodos de enseñanza y evaluación del

aprendizaje de los estudiantes, competencia en el manejo del aula, compromiso con el desarrollo profesional y cumplimiento de las responsabilidades profesionales (Departamento de Educación de Iowa, 2002). Sin embargo, los dos criterios restantes difieren de la definición de Danielson de lo que es una buena enseñanza. El primero se refiere a las estrategias utilizadas para impartir una enseñanza que cumpla con las múltiples necesidades de aprendizaje de los estudiantes. Danielson sostiene que la equidad está implícita en todo el *Marco*, particularmente en el caso de los dominios relacionados con la interacción con los estudiantes (“en un ambiente de respeto e interacción, *todos* los estudiantes se sienten valorados”), reconociendo, sin embargo, que “la conciencia de la adecuación a las diferentes etapas de desarrollo puede extenderse, incluyendo la sensibilidad a los estudiantes con necesidades especiales”. En contraste, el estándar de Iowa se centra explícitamente en la habilidad del maestro para aumentar la equidad, por ejemplo, aplicando estrategias que aborden todo el rango de niveles cognitivos o incorporando los conocimientos anteriores de los estudiantes, sus experiencias de vida y sus intereses en el proceso pedagógico. El segundo estándar de Iowa que difiere del *Marco* es la demostración por parte de los maestros de la capacidad para mejorar el rendimiento académico de los estudiantes. Por ejemplo, el maestro debe entregar evidencia del aprendizaje de los estudiantes a los estudiantes mismos, a las familias y al personal docente. Este punto es más conflictivo, no solo debido a que una multitud de otros factores puede influir en los indicadores de aprendizaje de los estudiantes, sino que además, y más importante, puede que no sea parte del rol del maestro demostrar a un tercero que su desempeño es la base del éxito académico de sus estudiantes. Los resultados de los estudiantes son utilizados algunas veces por los evaluadores como un instrumento para medir el desempeño de los maestros, pero el requerimiento de esta evidencia al maestro generalmente no es en sí mismo un criterio de buena práctica sujeto a evaluación.

#### **2.2.4 Instrumentos de recopilación de datos**

Si bien, en general, existe un amplio consenso con respecto a los criterios que determinan una buena enseñanza, la definición de los instrumentos para la recopilación de evidencia con respecto a la práctica actual de un docente es un aspecto mucho más conflictivo. Dado que la

forma de recopilar evidencia acerca de un docente en particular podría influir en los resultados de la evaluación, la selección de los instrumentos tiene una importancia crucial en el diseño e implementación de los sistemas de evaluación del desempeño docente.

**Observaciones en el aula.** Las observaciones en el aula son la fuente más común de evidencia utilizada en los países de la OCDE, ya sea en América (por ejemplo, Canadá, Chile, Estados Unidos), en Europa (por ejemplo, Dinamarca, Francia, Irlanda, España) o en la región de Asia-Pacífico (por ejemplo, Australia, Japón, Corea). Este proceso permite observar si el maestro adopta prácticas adecuadas en su lugar de trabajo más usual: el aula (UNESCO, 2007). Sin embargo, dependiendo del evaluador y del contexto, la utilidad y capacidad informativa de la evidencia recopilada podría diferir. Peterson (2000) explica que la observación de la experticia que exhibe un maestro en los contenidos que debe dominar ejerce un rol menor en algunas situaciones, pero es muy importante en otras.

**Entrevistas al docente.** Las entrevistas a los docentes podrían adoptar múltiples formas, con un grado de estructuración variable. En unos pocos casos, son útiles para formular juicios directos acerca de las competencias y habilidades de un maestro, pero son más adecuadas para fines de desarrollo profesional, al preguntarles a los maestros qué actividades de perfeccionamiento quisieran o necesitarían realizar. Por ejemplo, las escuelas inglesas que entregan formación profesional exhaustiva a los docentes utilizan entrevistas con respecto al manejo del desempeño para identificar las necesidades individuales de su personal docente (Ofsted, 2006). Sin embargo, la propensión de los docentes a revelar sus verdaderas deficiencias y temores durante las entrevistas depende de su confianza en el entrevistador y de sus percepciones de la posibilidad de recibir retroalimentación pertinente y constructiva derivada del proceso de evaluación.

**Portafolio preparado por el docente.** Los portafolios requieren que los docentes recopilen documentación acerca de su trabajo actual. Diferentes elementos pueden componer los portafolios preparados por los docentes: planes de las lecciones y materiales pedagógicos, grabaciones pedagógicas en video, muestras de trabajos realizados por los estudiantes y comen-

tarios con respecto a ejemplos de evaluaciones aplicadas a los estudiantes, cuestionarios autoadministrados por los docentes y hojas de reflexión. Beck, Livne y Bear (2005) recalcan que un dilema importante en el diseño de los portafolios es si el mismo es principalmente un medio para la evaluación docente o el desarrollo profesional docente y si estos dos objetivos son compatibles. Este punto se profundizará más adelante. Además, debido a que los portafolios constituyen una fuente de evidencia compleja, Wertzler y Strudler (2006), Strudler y Wertzler (2008), Jun *et al.* (2007) y Jacobs, Martin y Otieno (2008) señalan que la preparación de portafolios completos es particularmente útil en la evaluación de los estudiantes de pedagogía o los maestros principiantes. Los programas de formación docente también podrían beneficiarse a partir de los resultados obtenidos por los maestros principiantes en las evaluaciones de sus portafolios (Jacobs, Martin y Otieno, 2008).

**Indicadores de rendimiento escolar.** Los indicadores de rendimiento escolar no suelen ser utilizados como fuentes de evidencia para la evaluación docente en los países de la OCDE (OCDE, 2005; UNESCO, 2007). Los resultados de los estudiantes en rendimiento académico podrían reflejar el desempeño docente, especialmente cuando se los mide en términos de mejoramientos de valor agregado más que en términos absolutos, es decir, después de controlar los resultados anteriores de estudiantes individuales que recibieron enseñanza del maestro (Braun, 2005). Por ejemplo, la Evaluación del Desempeño Docente de California (*Californian Teacher Performance Assessment*) mide los mejoramientos en el aprendizaje de los estudiantes en relación con los estándares de los distritos con el fin de recomendar a los maestros para una certificación. Sin embargo, el aprendizaje de los estudiantes rara vez es utilizado como una medida del desempeño de los maestros en los esquemas existentes, ya sea debido a que no existen pruebas estandarizadas regulares que permitan comparaciones viables o debido a que enfrentan un fuerte rechazo de los docentes y los especialistas, que consideran que este instrumento es defectuoso, ineficaz o poco equitativo (Weingarten, 2007). Las ventajas y desventajas de este tipo de medición directa del 'desempeño puro' se analizarán más adelante.

**Pruebas a los docentes.** Excepcionalmente, el conocimiento curricular y las habilidades peda-

gógicas de los maestros se evalúan mediante pruebas escritas. Es lo que ocurre en el caso de los nuevos maestros en Chile (Ávalos y Assael, 2006) o los maestros que postulan a un ascenso en México (OCDE, 2005).

**Cuestionarios y encuestas.** Los cuestionarios con respecto a la práctica de los docentes podrían ser respondidos por el director del establecimiento, por los padres o por los estudiantes, es decir, por quienes podrían dar testimonio de la calidad de la enseñanza como resultado de una interacción permanente con el maestro y no solo durante el proceso de evaluación (Peterson, 2000; Peterson *et al.*, 2000, 2005; Jacob y Lefgren, 2005b). En consecuencia, esta categoría precisa de cuestionarios y encuestas excluye los informes de evaluadores producto de observaciones del aula o las entrevistas al maestro; está restringida a cuestionarios como fuentes de evidencia *per se*. Las encuestas a los estudiantes son utilizadas como herramientas de evaluación docente en México, la República Eslovaca, España o Suecia, generalmente en el caso de los maestros que postulan a un ascenso; según nuestros conocimientos, no se aplican en los esquemas de evaluación docente obligatorios. Si bien la utilización de estos instrumentos puede aportar alguna información interesante, es preciso tener en mente que los evaluadores no son expertos en docencia y no valoran necesariamente las mismas cualidades que aquellas que se supone que favorecen el aprendizaje de los estudiantes (Peterson *et al.*, 2000, 2003; Jacob y Lefgren, 2005b). Las investigaciones con respecto al uso o la confiabilidad de dichos procedimientos lamentablemente siguen siendo muy escasas.

### 2.2.5 Evaluadores

**Revisión interna.** En la mayoría de los países, las evaluaciones docentes contemplan la participación del director del establecimiento escolar u otro personal directivo (Peterson, 2000; OCDE, 2005, 2008; UNESCO, 2007). Sin embargo, el compromiso del personal directivo de los establecimientos escolares en el proceso de evaluación difiere en los distintos países. En 2003, el 100% de los estudiantes norteamericanos estaban matriculados en escuelas secundarias cuyos directores informaban que habían realizado observaciones de aula durante el año anterior, mientras que esto solo era válido para el 5% de los estudiantes de Portugal, con un promedio de

60% para la OCDE (OCDE, 2008). En los distintos países, es variable el tiempo y la capacidad que tienen los directivos de los establecimientos escolares para asumir esta importante responsabilidad (Marshall, 2005; Jacob y Lefgren, 2005a, 2008). Las ventajas y desventajas derivadas del hecho que la evaluación sea realizada por los directores en lugar de otros evaluadores se examinarán más adelante.

**Revisión externa.** Algunos países han implementado esquemas de evaluación en los cuales los docentes son evaluados por pares o por maestros competentes, ya sea en forma exclusiva (Irlanda) o como parte de un panel que incluye al director del establecimiento (Francia). Por una parte, los ‘pares’ son otros docentes que tienen un nivel equivalente en términos de asignación, formación, experiencia, perspectiva e información acerca de las condiciones en las cuales se desenvuelve la práctica sujeta a revisión, pero no deberían ni enseñar en la misma escuela que el maestro que está siendo evaluado ni estar social o políticamente relacionados con este (Peterson, 2000). Por otra parte, los ‘maestros’ competentes son reconocidos como maestros que tienen un conocimiento profundo de las materias y experticia pedagógica, como profesionales altamente competentes y exitosos, capaces de guiar y apoyar a otros en el proceso de enseñanza (MCEETYA, 2003). Ambos tipos de evaluadores tienen ventajas relativas: los primeros participan en el proceso en una condición de igualdad con los docentes evaluados, mientras que los segundos lo hacen desde la perspectiva de la experticia.

**Autoevaluación.** La participación de los docentes en ‘evaluaciones de empoderamiento’ es esencial tanto para obtener la aceptación del proceso de evaluación por parte de los maestros como para mejorar el desempeño docente (Peterson, 2000; Kennedy, 2005). Los portafolios son instrumentos particularmente adecuados para la auto-reflexión docente, debido a que la adecuada decisión tomada por el maestro de incluir determinados componentes (planes de las lecciones, grabaciones de lecciones en video, muestras de trabajos realizados por los estudiantes, comentarios narrativos) en lugar de otros, constituye un juicio que requiere determinar qué características de un componente son superiores a otras (Danielson, 1996, 2007; Darling, 2001; Mansvelder-Longayroux *et al.*, 2007). En combinación con una revisión realizada por

otro evaluador, los documentos preparados por el maestro podrían utilizarse con un propósito sumativo. Sin embargo, el propósito formativo es predominante, puesto que el proceso de reflexión permite que el maestro tome conciencia de sus propias fortalezas y deficiencias e identifique sus necesidades de perfeccionamiento, desarrollo profesional o *coaching*.

**Padres.** Los padres generalmente desempeñan una función indirecta en el proceso de evaluación cuando los informes de los directores incluyen sus reclamos acerca de un maestro en particular o, por el contrario, sus solicitudes con respecto a un maestro en particular. Es menos frecuente que actúen como evaluadores directos, como, por ejemplo, a través de cuestionarios. La escasa evidencia actual en esta área muestra que las características de los maestros valoradas por los padres están sorprendentemente alejadas del rendimiento académico de los estudiantes: ‘la capacidad del maestro de promover la satisfacción de los estudiantes’ (Jacob y Lefgren, 2005b), ‘el trato humano a los estudiantes’, ‘el apoyo al aprendizaje de los estudiantes’ y ‘la comunicación y colaboración efectiva con los padres’ (Peterson *et al.*, 2003). Incluso si su perspectiva pudiera tomarse en consideración, su distancia de los estándares profesionales docentes, su falta de conocimiento acerca de lo que ocurre en las aulas y su compromiso emocional sugieren que sus valoraciones están lejos de ser apropiadas para un esquema de evaluación docente integral.

**Estudiantes.** Los estudiantes son rara vez consultados como evaluadores. México, España y Suecia utilizan encuestas a los estudiantes, pero generalmente solo en algunos grados de la enseñanza o en casos especiales de evaluación docente (por ejemplo, evaluaciones voluntarias para un ascenso o un procedimiento relacionado con un reclamo). Los estudios con respecto a evaluaciones de los docentes realizadas por estudiantes en los niveles de la educación primaria y secundaria son extremadamente escasos. Peterson *et al.* (2000) señalan que los estudiantes responden en forma directa y confiable con respecto a la calidad de los maestros si las preguntas están formuladas de manera simple y pertinente. Propusieron tres conjuntos de preguntas que, según ellos, funcionan bien, para estudiantes de la enseñanza preescolar, primaria y secundaria. Sin embargo, consideran a los estudiantes como “clientes”, pero este punto es muy

cuestionable. En efecto, los estudiantes no pagan directamente por los servicios educacionales (mientras que sus padres pueden que paguen, pero no son los consumidores de los servicios educacionales) y, lo que es más importante, los estudiantes son miembros matriculados involuntariamente: no tienen la libertad de abandonar la organización, de elegir su escuela o sus maestros, ni de influir en lo que se les enseña o cómo se les enseña (Greenfield, 1995).

### 2.3 VÍNCULOS CON EL RECONOCIMIENTO Y LAS RETRIBUCIONES

El proceso de evaluación podría estar vinculado a mecanismos de reconocimiento y retribuciones. En la mayoría de los países de la OCDE existe una escala salarial única para los maestros y pocos incentivos formales y reconocimientos por la buena práctica. Esto plantea inquietudes con respecto al atractivo de la docencia como una elección de carrera profesional y al mantenimiento de la motivación de los maestros durante su carrera. El Departamento de Educación de Australia sostiene que, si bien las personas que han escogido la docencia como carrera están principalmente motivadas por las recompensas 'intrínsecas', factores extrínsecos tales como la remuneración son los factores más significativos que influyen en el hecho de que las personas no escojan la docencia como carrera —especialmente en el caso de los posibles postulantes de alta calidad— y que influyen en el hecho de que los maestros abandonen la profesión (DEST, 2007).

Heneman *et al.* (2006) sostienen que los sistemas de evaluación docente basados en estándares deben utilizarse como base para la remuneración basada en los conocimientos y las habilidades. Apoyan una estrategia de incentivos que requiere el diseño e implementación de sistemas de remuneración docente alternativos que difieren de la escala salarial única. Esta nueva estrategia, aplicada actualmente por varios estados norteamericanos (Peterson, 2006), vincula la remuneración con combinaciones de evaluaciones del desempeño docente, la adquisición de nuevos conocimientos y habilidades, y el mejoramiento de los resultados obtenidos por los estudiantes en las pruebas estandarizadas. Pocos países europeos entregan un aumento directo de la remuneración sobre la base salarial como retribución por un buen desempeño. Por ejemplo, Rumania ha establecido un sistema en

el cual los mejores maestros pueden competir por un aumento temporal de la remuneración, del 15% durante un año al 20% durante cuatro años (UNESCO, 2007).

Otros países no relacionan directamente los resultados de las evaluaciones docentes con la remuneración de los maestros, pero los vinculan a la progresión en la carrera. Francia, Alemania, Grecia, Polonia, Portugal y el Reino Unido se encuentran entre estos países. Los maestros ingleses que cumplen con los estándares para las categorías de Maestros Post-Umbrales, Maestros Excelentes y Maestros con Habilidades Avanzadas, también acceden a la escala de remuneraciones correspondiente (TDA, 2007b). En contraste, en una serie de países de la OCDE, los resultados continuamente negativos en las evaluaciones están vinculados a postergaciones del ascenso (OCDE, 2005). Cabe señalar que la vinculación del proceso de evaluación a incrementos salariales a través del escalafón podría no ser suficiente en países en los cuales los niveles superiores del escalafón se alcanzan en etapas tempranas de la carrera docente. Por ejemplo, en Australia, el incentivo de incrementos salariales vinculados a una revisión del desempeño no es aplicable a la mayoría de los maestros, puesto que ya se encuentran en el extremo superior de la escala de remuneraciones después de 10-12 años de carrera docente y, por lo tanto, no parece ser particularmente eficaz (Kleinhenz e Ingvarson, 2004).

Finalmente, algunos países vinculan las evaluaciones docentes con oportunidades de ascensos verticales a cargos de liderazgo escolar. En España, una de las condiciones para ser elegido como jefe del Consejo Escolar es aprobar el proceso de evaluación docente (UNESCO, 2007). Asimismo, en los años posteriores a la etapa de progresión de la carrera de 10<sup>o</sup> o 12<sup>o</sup> año en Australia, algunos maestros muy competentes son ascendidos a cargos administrativos que pueden llegar hasta el puesto de director (Kleinhenz y Ingvarson, 2004). Sin embargo, la práctica de vincular el desempeño docente destacado a ascensos verticales puede ser criticada, por dos razones principales. Por una parte, un buen maestro no es necesariamente un buen administrador ni un buen directivo. Por otra parte, esta práctica podría tener efectos adversos en su intento por obtener el reconocimiento de la profesión docente; paradójicamente, los mejores maestros son retribuidos por no seguir ha-

ciendo lo que hacen mejor: enseñar (UNESCO, 2007). En efecto, los maestros destacados que escogen seguir enseñando son considerados como “criaturas oscuras que ocupan el submundo del aula”, cuyos conocimientos y habilidades no suelen ser reconocidos. Adoptando la dirección opuesta, algunas escuelas australianas han optado por implementar una categoría docente entre el nivel superior de la escala salarial automática y las categorías administrativas, con el fin de retribuir a los maestros que optan por permanecer en el aula en lugar de pasar a la administración (Kleinhenz y Ingvarson, 2004).

#### **2.4 VÍNCULOS CON LAS OPORTUNIDADES DE DESARROLLO PROFESIONAL Y LAS PRIORIDADES MÁS GENERALES DE LAS ESCUELAS**

Pocos países vinculan la evaluación del desempeño con la formación profesional continua. Sin embargo, es esencial la existencia de una cadena lógica entre la evaluación del desempeño y las oportunidades de formación profesional continua para mejorar la práctica docente (Ofsted, 2006). La identificación de las fortalezas y deficiencias de cada uno de los maestros es importante para escoger a partir de una amplia gama de posibles actividades de desarrollo profesional aquellas que cumplen las necesidades particulares de los maestros en función de cada una de las prioridades del plan de mejoramiento del establecimiento escolar correspondiente.

En el Reino Unido se implementaron nuevos planes e iniciativas en esta dirección. Desde 2005-2006, la Agencia de Capacitación y Desarrollo para las Escuelas (*Training and Development Agency for Schools* - TDA) está a cargo de coordinar el desarrollo profesional de todo el personal de los establecimientos escolares ingleses. En septiembre de 2007, se implementaron nuevos estándares docentes con el fin de establecer un marco para la evaluación docente de acuerdo con las políticas escolares más generales. Estos estándares dan énfasis al vínculo entre lo que se espera de un ‘buen’ maestro en cada etapa de la carrera profesional, por una parte, y las oportunidades de perfeccionamiento con miras a la siguiente etapa de la carrera, por otra. “El marco ofrece un telón de fondo para los debates con respecto a la manera de considerar el desempeño de un maestro en relación con la etapa actual de su carrera profesional y la etapa a la que se está aproximando. Los estándares

pertinentes deben ser visualizados en su conjunto con el fin de ayudar a los maestros a identificar sus áreas de fortaleza y las áreas que requieren perfeccionamiento. Un maestro que aspira a acceder a una etapa superior de la carrera profesional deberá reflexionar y analizar la mejor forma de planificar su futuro desarrollo profesional de modo de poder trabajar en pos del cumplimiento de los estándares y el manejo del desempeño proporcionaría evidencia para la futura postulación del maestro” (TDA, 2007b). Las escuelas que asocian las necesidades individuales identificadas con las prioridades propias de la escuela y que también logran desarrollar las correspondientes actividades de formación profesional continua, tienen una alta probabilidad de lograr un buen desempeño (Ofsted, 2006).

Sin embargo, todavía queda mucho por hacer en esta área. Margo *et al.* (2008) recalcan que actualmente existen muchos problemas que obstaculizan el logro de una fuerza laboral docente plenamente eficiente en Inglaterra, entre los cuales se encuentra una ‘calidad inconsistente de la formación docente’ y un ‘desarrollo profesional inadecuado’. Para superar estos problemas, recomiendan reforzar nuevamente el vínculo entre la formación profesional continua (FPC) y el proceso de evaluación, a través de evaluaciones más frecuentes, requisitos integrados de FPC y la obligación de que los maestros cuya evaluación indique un bajo desempeño accedan a actividades de perfeccionamiento apropiadas antes de reingresar a la docencia.

La percepción de los procedimientos de evaluación como base para el mejoramiento de la práctica futura es crucial para implementar un sistema en el cual todos los docentes se sientan comprometidos con la evaluación y con las oportunidades de desarrollo profesional pertinentes, independientemente de su nivel de desempeño actual. Evidentemente, los procedimientos de evaluación son necesarios para adoptar las medidas adecuadas ante los docentes ineficientes y garantizar que adopten las prácticas apropiadas. Sin embargo, sin un vínculo con las oportunidades de desarrollo profesional, el proceso de evaluación no es suficiente para mejorar el desempeño docente y, como resultado, suele convertirse en un ejercicio sinsentido enfrentado con desconfianza –o, en el mejor de los casos, apatía– por los maestros que están siendo evaluados (Danielson, 2001;

Milanowski y Kimball, 2003; Margo *et al.* 2008; Pochard, 2008). Con respecto al sistema francés, Pochard (2008) sostiene que lamentablemente los programas de desarrollo profesional no están diseñados como para responder a las necesidades de perfeccionamiento claramente identificadas tanto por el docente como por la institución. Se sostiene que la evaluación por sí sola no es suficiente como para implementar los cambios necesarios para promover mejoramientos de la eficacia y la equidad del sistema educacional. También se plantea que toda evaluación que identifique las disfunciones existentes en un establecimiento escolar debería traducirse en el diseño de un nuevo plan educacional apoyado por un equipo externo.

### **3. VENTAJAS Y DESVENTAJAS DE LOS DIFERENTES ENFOQUES CON RESPECTO A LA EVALUACIÓN DOCENTE**

En esta sección se analiza una serie de aspectos conflictivos relacionados con el diseño y la implementación de los sistemas de evaluación docente. El primer aspecto controvertido es si deberían utilizarse los indicadores de rendimiento escolar como una medida del desempeño de los docentes. El segundo punto de debate ilustra que las ventajas y desventajas de los diferentes métodos generalmente están relacionadas con los propósitos planteados para la evaluación docente y que, dadas las usuales restricciones de recursos, es inevitable realizar compensaciones entre los argumentos en favor de enfoques sumativos y aquellos en favor de sistemas formativos. Finalmente, los argumentos en favor o en contra de los diferentes enfoques reflejan las diferentes perspectivas de los actores, las que se traducen en dificultades de implementación.

#### **3.1 VENTAJAS Y DESVENTAJAS DEL USO DE LOS INDICADORES DE RENDIMIENTO ESCOLAR COMO MEDIDA DEL DESEMPEÑO DOCENTE**

Los indicadores de rendimiento escolar constituyen una atractiva medida para evaluar el desempeño docente, puesto que el objetivo final de la enseñanza es el mejoramiento del aprendizaje de los estudiantes. No es de sorprender que gran parte de las investigaciones se hayan centrado en el uso del rendimiento académico de los estudiantes medido a través de las pruebas estandarizadas para evaluar a los docentes.

Por ejemplo, Leigh (2007) examinó recientemente los resultados de las pruebas de alfabetización y conocimientos básicos de matemáticas de tres cohortes de estudiantes y concluyó que las variaciones en las posiciones relativas de las clases de estudiantes proporcionaban una base para la identificación de los maestros eficientes e ineficientes. Braun (2005) sostiene que la consideración de los resultados de los estudiantes es un enfoque prometedor por dos razones: en primer lugar, traslada el debate acerca de la calidad de los docentes hacia el aprendizaje de los estudiantes como el principal objetivo de la enseñanza y, en segundo lugar, incorpora una medida cuantitativa –y, por ende, objetiva y equitativa– del desempeño de los docentes. A este respecto, el desarrollo de modelos de “valor agregado” representa un significativo avance en relación con los métodos basados en la proporción absoluta de estudiantes que logran un nivel de rendimiento determinado. Los modelos basados en el “valor agregado” tienen por objetivo controlar los puntajes anteriores obtenidos en las pruebas por cada uno de los estudiantes y, por lo tanto, tienen la potencialidad de identificar el aporte de un maestro particular al rendimiento académico de los estudiantes.

En Florida, el esquema denominado “Los Maestros Especiales reciben una Retribución” (*Special Teachers are Rewarded* - STAR) vincula la remuneración o las bonificaciones para maestros particulares a medidas de valor agregado del aprendizaje de los estudiantes (Ingvarson, Kleinhenz y Wilkinson, 2007). Sin embargo, este tipo de vínculo entre una medida de desempeño directa y la remuneración sigue siendo extremadamente escaso, dados los numerosos desafíos estadísticos y teóricos asociados al uso de estos métodos. En efecto, Braun (2005) recalca el marcado contraste entre el entusiasmo de quienes quisieran utilizar tales mediciones, principalmente las autoridades a cargo de las políticas, y las reservas expresadas por los investigadores que han estudiado sus características técnicas.

El uso del rendimiento académico de los estudiantes en las pruebas estandarizadas para evaluar el desempeño de los docentes plantea numerosos desafíos estadísticos. La mayoría de los autores (Lockwood, Louis y McCaffrey, 2002; Kupermintz, 2003; Braun, 2005; Aaronson, Barrow y Sander, 2007; Goe, 2007) no están convencidos de que la generación actual de modelos de valor agregado sea lo suficientemente válida y confia-

ble como para ser utilizada para evaluar en forma equitativa la eficiencia de maestros particulares. Las limitaciones estadísticas están relacionadas, en primer lugar, con la notoria ausencia de datos confiables, principalmente debido al hecho de que los estudiantes no suelen rendir pruebas estandarizadas con una frecuencia anual. Rowley e Ingvarson (2007) critican la metodología de Leigh (2007), que consiste en crear un puntaje de prueba hipotético en el año con datos faltantes en el punto medio de dos resultados disponibles, argumentando que no permite atribuir en forma equitativa el éxito de los estudiantes a los diferentes maestros involucrados. En segundo lugar, cuando se dispone de datos, las variaciones del muestreo pueden provocar imprecisiones en las medidas de los puntajes de las pruebas. Este problema es particularmente notorio en las escuelas básicas, donde el limitado número de estudiantes por clase genera grandes idiosincrasias de la muestra particular de estudiantes que está siendo evaluada (Kane y Staiger, 2002).

Las principales críticas metodológicas señalan que los modelos basados en el valor agregado, cualquiera sea su grado de sofisticación, no pueden ni integrar cabalmente todos los factores que influyen en los resultados obtenidos por los estudiantes en las pruebas –de naturaleza cualitativa– ni reflejar todos los resultados de aprendizaje de los estudiantes. Los antecedentes y el apoyo de las familias, la asistencia a la escuela, el clima entre pares y en el aula, las políticas de las escuelas, la disponibilidad de materiales apropiados y los efectos de los niños influyen en el aprendizaje de los estudiantes (CAESL, 2004; Ingvarson, Kleinhenz y Wilkinson, 2007; Goe, 2007, Weingarten, 2007). Factores específicos que intervienen en el momento de la prueba –“un perro que ladra en el patio, una temporada con una alta incidencia de resfríos, un estudiante con mal comportamiento en una clase”– también pueden afectar los resultados de un estudiante independientemente del aporte de su maestro (Kane y Staiger, 2002). Además, es probable que los buenos maestros tengan un impacto en el rendimiento de los niños durante varios años después de haberles enseñado y, a la inversa, después de varios años de estar sometidos a maestros ineficientes, puede que los estudiantes nunca sean capaces de ponerse al día académicamente. Estos ‘efectos acumulativos’ de los docentes no pueden medirse con precisión en puntos discretos en el tiempo (Hanushek, 1986; Sanders y Ri-

vers, 1996; CAESL, 2004). Finalmente, el impacto de la enseñanza en los estudiantes no está restringido a las áreas evaluadas a través de las pruebas estandarizadas, generalmente limitadas a lectura y matemáticas, sino que también incluyen la transferencia de habilidades psicológicas, cívicas y de aprendizaje para toda la vida (Margo *et al.*, 2008). Xin, Xu y Tatsuoka (2004) intentaron descomponer los puntajes de pruebas estandarizadas en diversas categorías de habilidades cognitivas en cuatro países (Japón, Corea, Holanda y Estados Unidos) y encontraron que los atributos de los maestros considerados en las decisiones con respecto a las remuneraciones no tienen un impacto positivo consistente sobre ningún tipo de habilidad cognitiva, a pesar de haber controlado los antecedentes tanto individuales como familiares. Estas son fuentes de escepticismo con respecto al uso de estos métodos estadísticos.

También es necesario considerar las limitaciones teóricas. En primer lugar, una correlación estadística no es una relación causal: el hecho de que los docentes sean importantes para el aprendizaje de los estudiantes no indica necesariamente que el aprendizaje de los estudiantes sea el resultado de una buena enseñanza. En segundo lugar, las pruebas estandarizadas utilizadas para evaluar a los estudiantes no están específicamente diseñadas para efectos de evaluar a los docentes. Siguiendo a Popham (1997), Goe (2007) sostiene que no están diseñadas para ser particularmente sensibles a pequeñas variaciones en la pedagogía o para determinar los aportes de los docentes al aprendizaje de los estudiantes. En consecuencia, no ofrecen una base sólida que permita responsabilizar a los docentes por los resultados de su desempeño. En tercer lugar, el uso de los puntajes obtenidos por los estudiantes en las pruebas para evaluar a los docentes podría inducir distorsiones y constricciones no esperadas en el comportamiento de los docentes, centrándose solo en el rendimiento de los estudiantes en las pruebas estandarizadas. Los esquemas de incentivos de altas consecuencias (*high-stakes*) basados en las pruebas estandarizadas pueden incitar a los docentes a concentrarse exclusivamente en las áreas pedagógicas evaluadas en las pruebas, reduciendo así el currículo a las habilidades básicas generalmente evaluadas (Jacob y Lefgren, 2005; Weingarten, 2007), incitar a los maestros a concentrarse en los estudiantes específicos que están cerca de la nota

de aprobación a expensas de los niños que están más atrasados o adelantados (Weingarten, 2007) e incluso provocar graves casos de engaño por parte de los docentes en las pruebas estandarizadas (Jacob y Levitt, 2003; Jacob, 2005). Por otra parte, puede que los resultados de las pruebas permitan identificar a los maestros que son ineficientes o requieren perfeccionamiento, pero no permiten ni discriminar en forma justa entre la amplia gama de maestros eficientes ni identificar cuáles son las actividades de perfeccionamiento que deberían implementarse con el fin de mejorar el desempeño de los maestros ineficientes (Braun, 2005). Finalmente, estos sistemas podrían traducirse en que los maestros fuesen responsabilizados por todo el desempeño de los estudiantes mientras que, por el contrario, se debería reconocer que la enseñanza exitosa es una responsabilidad compartida entre los gobiernos, las escuelas y la profesión docente (Ingvarson *et al.*, 2007).

Como consecuencia, a pesar del atractivo de la idea, existen numerosas advertencias contra el uso de los puntajes obtenidos por los estudiantes en las pruebas para evaluar a los docentes. En particular, existe un amplio consenso en la literatura en torno a dos direcciones específicas: los resultados de los estudiantes no deberían utilizarse como la *única* medida del desempeño de los docentes y no deberían utilizarse ingenuamente para la toma de decisiones relativas a la carrera profesional de un docente, incluyendo el vínculo con la remuneración, debido a que esto incorpora un considerable riesgo de castigar o retribuir a los docentes por resultados que están fuera de su control (Kane y Staiger, 2002; Kupermintz, 2002; McCaffrey *et al.*, 2003; CAESL, 2004; Raudenbush, 2004; Braun, 2005; Ingvarson, Kleinhenz y Wilkinson, 2007; Rowley e Ingvarson, 2007). Estas objeciones de los docentes y los investigadores se han materializado, por ejemplo, en la decisión de la legislatura del Estado de Nueva York de prohibir el uso de los resultados obtenidos por los estudiantes en las pruebas para evaluar a los maestros en abril de 2008.

### **3.2 DISEÑO DE UN CONJUNTO COHERENTE DE MÉTODOS E INSTRUMENTOS ALINEADO CON EL PROPÓSITO ENFATIZADO DE EVALUACIÓN DOCENTE**

Fenstermacher y Richardson (2005) distinguen dos enfoques para determinar la calidad de los

docentes: la “enseñanza exitosa” es una medida del desempeño puro, en tanto que la “buena enseñanza” se centra en la calidad de las oportunidades de aprendizaje de los estudiantes ofrecidas en las aulas en relación con los estándares docentes. Dado que no existe una relación directa, confiable y segura entre la calidad de los docentes y el rendimiento de sus estudiantes en las pruebas estandarizadas, existe un amplio consenso en que es preferible evaluar a los maestros por su práctica (una “buena enseñanza”) en lugar de por su desempeño (una “enseñanza exitosa”) (Ingvarson *et al.*, 2007). Como resultado, la recopilación de *múltiples* fuentes de evidencia acerca de la práctica docente satisface las necesidades de rigurosidad y equidad del proceso de evaluación, tomando en consideración la complejidad de lo que un ‘buen’ maestro debería saber y ser capaz de hacer (Danielson, 1996, 2007; Peterson, 2000).

Sin embargo, mientras que la multiplicación de instrumentos y evaluadores con mayor probabilidad entregará una base sólida para evaluar a los docentes, la limitación de los recursos hace inevitables las compensaciones. Los esquemas de evaluación docente integrales implican mayores costos directos e indirectos en cada etapa del proceso: el alcance de acuerdos sobre el diseño del sistema requiere tiempo para debates y consultas con todos los actores (Ávalos y Assael, 2006); la capacitación de los evaluadores es onerosa (Danielson, 1996, 2007); la conducción de procesos de evaluación agrega una carga de trabajo adicional tanto para los docentes como para los evaluadores, a menos que se realicen compensaciones reduciendo la carga de trabajo derivada de otras responsabilidades (Heneman *et al.*, 2006); la armonización de las reformas escolares más generales, tales como las oportunidades de desarrollo profesional, requiere más recursos educacionales (Heneman *et al.*, 2007; Margo *et al.*, 2008). Por estas razones, a menudo los países han decidido implementar solo algunos de los aspectos de los esquemas de evaluación docente. Sin embargo, la implementación de compensaciones entre los diferentes métodos no es simple, puesto que las ventajas de instrumentos o métodos particulares para efectos sumativos suelen constituir desventajas para efectos formativos y viceversa.

En lo que respecta a las fuentes de evidencia, los instrumentos tales como los indicadores de rendimiento escolar, las pruebas a los maestros,

los cuestionarios y encuestas respondidas por los padres y los estudiantes, y las observaciones en el aula, tienen una naturaleza más sumativa, mientras que las entrevistas con el docente y la documentación preparada por el docente generalmente son más útiles para efectos formativos. Los mecanismos de responsabilidad por los resultados requieren calificar cuantitativa y objetivamente a los maestros de acuerdo a un marco único compuesto por unos pocos aspectos profesionales (Nabors Oláh, Lawrence y Riggan, 2008). Esto permite comparar a los docentes de acuerdo a estándares idénticos bien definidos y totalizar fácilmente los puntajes obtenidos en los diferentes criterios. Un índice global de desempeño docente obtenido de esta manera podría ofrecer una escala equitativa para reconocer y retribuir el trabajo de los maestros. En contraste, cuando el propósito es ayudar a los maestros a mejorar sus prácticas y ofrecerles oportunidades de desarrollo profesional, son preferibles los instrumentos y criterios cualitativos especialmente adaptados. Para efectos formativos, la recopilación de evidencia adaptada es más apropiada que un conjunto de estándares que se ajusten a todas las situaciones posibles. Debe permitir tanto la identificación de los dominios de fortalezas y deficiencias en la docencia como la entrega al maestro de una retroalimentación constructiva que incluya las posibles maneras de mejorar su desempeño, de acuerdo al nivel de experiencia del maestro y al contexto de la escuela.

Los portafolios son particularmente interesantes en la medida que contienen componentes de trabajo docente que pueden combinarse de diferentes maneras de acuerdo al propósito enfatizado. Por una parte, Klecker (2000), Campbell *et al.* (2000) y Tucker *et al.* (2002) sostienen que los portafolios entregan información evaluativa que permite hacer responsables a los maestros por el cumplimiento de los estándares educacionales. Por otra parte, Darling (2001) sostiene que el desarrollo profesional de los docentes debe tener prioridad en el diseño de los portafolios y que la 'reflexión narrativa' es la mejor manera de estimular dicho desarrollo. Beck, Livne y Bear (2005) observan que los portafolios formativos centrados en el desarrollo profesional de los docentes apoyan de mejor manera los resultados profesionales y, en consecuencia, sostienen que los portafolios no deberían ser utilizados para la evaluación sumativa de los docentes. Otra interrogante es si los portafolios electrónicos son más efica-

ces y justos que los portafolios impresos. Los portafolios electrónicos facilitan el acceso constante a las muestras de trabajos y motivan la retroalimentación de los evaluadores, permiten captar el reflejo inmediato del desarrollo profesional de los docentes y permiten una fácil presentación de múltiples puntos de datos (Wetzel y Strudler, 2006; Jun *et al.*, 2007), mientras que presentan una desventaja en comparación con el portafolio impreso en el hecho de que un maestro con mayores habilidades tecnológicas se encuentra en ventaja, incluso si sus capacidades docentes son precisamente las mismas (Wetzel y Strudler, 2006). Strudler y Wetzel (2008) también destacan la significativa cantidad de tiempo y esfuerzo empleados en la creación y revisión de los portafolios y la posible falta de compatibilidad con las creencias, valores e intereses de las instituciones.

En lo que respecta a los evaluadores, existe un amplio consenso en el hecho de que los directores de las escuelas, los pares y los expertos, los padres y los estudiantes no valoran las mismas capacidades y conocimientos docentes, no se refieren al mismo conjunto de evidencia y tienen diferentes percepciones y grados de objetividad. Por consiguiente, es frecuente que la participación de *múltiples* evaluadores sea visualizada como una clave para las prácticas exitosas; por lo menos más de una persona debería estar involucrada en la calificación de la calidad y el desempeño de los docentes (Peterson, 2000; Stronge y Tucker, 2003). Danielson y McGreal (2000) explican que los 'sistemas de evaluación en 360 grados', que incorporan la participación de muchos tipos de evaluadores, apoyan la idea de que la competencia de un maestro podría considerarse desde varias perspectivas diferentes y que debería ser ejemplar (o por lo menos suficiente) desde todos estos diferentes ángulos.

La literatura distingue las ventajas y desventajas relativas de la inclusión de diferentes tipos de evaluadores de acuerdo a la naturaleza del proceso de evaluación. Los directores parecen ser particularmente eficaces en identificar a los mejores y a los peores docentes (aquellos que se encuentran en el 10-20% superior e inferior de la distribución) (Jacob y Lefgren, 2005a, 2008) y su función de supervisión y liderazgo los coloca *a priori* en una posición como para realizar evaluaciones de los docentes. Esto sugiere que pueden ser evaluadores relevantes para las evaluaciones sumativas. Sin embargo, la evidencia también

sugiere que los directores muestran muy pocas habilidades para distinguir entre los maestros que se encuentran en la amplia mayoría de la distribución, discriminan sistemáticamente basándose en algunas características de los maestros (por ejemplo, el género, la edad, la educación) y suelen mostrarse influidos por una serie de factores afectivos o no relacionados con el desempeño, tales como la afinidad con el subordinado o la primera impresión que tuvieron del maestro en particular que está siendo evaluado (Lefkowitz, 2000; Bolino y Turnley, 2003; Jacob y Lefgren, 2005a, 2008). Levin (2003) y MacLeod (2003) muestran que la exigencia e indulgencia del director en las evaluaciones del desempeño docente, que se encuentran frecuentemente en la práctica, son características del contrato óptimo entre un director neutro ante el riesgo y un agente con aversión al riesgo cuando las retribuciones se basan en una evaluación subjetiva del desempeño. Finalmente, es difícil considerar a los directores como jueces imparciales debido a que se encuentran en contacto diario con los maestros evaluados. Estos argumentos muestran que un sistema de evaluación basado en los directores podría ser útil para decisiones de despido, pero podría no ser lo suficientemente riguroso como para estar vinculado a incrementos de las remuneraciones. En contraste, los revisores externos evalúan a los maestros en relación con marcos de estándares profesionales, conocen las especificidades de contenidos y habilidades para cada área pedagógica, pero son relativamente menos capaces de adaptar el proceso al contexto, los problemas y los valores del establecimiento escolar en cuestión (Anderson y Pellicer, 2001).

En relación con el propósito formativo, también existen debates acerca de quién está en posición de definir con precisión las necesidades de perfeccionamiento de los docentes y entregar la retroalimentación más constructiva. Los pares y colegas que tienen las mismas características e imparten la misma materia al mismo grado de estudiantes exhiben la mayor probabilidad de ganarse la confianza del maestro evaluado. Por lo tanto, el maestro podría comprometerse con mayor facilidad en una auto-reflexión acerca de sus prácticas y expresar sus sentimientos y preocupaciones durante las entrevistas sin el miedo a posibles sanciones. Los pares también pueden entregar retroalimentación cualitativa basada en su propia experiencia en el área pedagógica pertinente.

Sin embargo, los directores son esenciales para vincular las necesidades de perfeccionamiento detectadas en el maestro a las oportunidades de formación profesional continua de acuerdo a los objetivos del establecimiento escolar y otras necesidades de personal docente. También exhiben una mayor probabilidad de entregar retroalimentación informal permanente al maestro durante todo el año y no solo durante el proceso de evaluación formal. En un contexto más general, son esenciales para convertir el mejoramiento del desempeño en un imperativo estratégico y ayudar a considerar la evaluación docente como indispensable para el maestro y para las políticas escolares más amplias (Heneman *et al.*, 2007; Robinson, 2007; OCDE, 2008).

### **3.3 DIFICULTADES EN LA IMPLEMENTACIÓN DE LOS ESQUEMAS DE EVALUACIÓN DEL DESEMPEÑO DOCENTE**

#### **3.3.1 Conflictos de interés entre los diferentes actores**

La selección de los métodos de evaluación se vuelve aún más compleja debido a las significativas divergencias entre las perspectivas e intereses de los diferentes actores. La importancia relativa de los propósitos sumativos y formativos es particularmente conflictiva. Por una parte, las autoridades encargadas de la formulación de las políticas y los padres tienden a valorar la garantía de calidad y la responsabilidad por los resultados. “Sostienen que las escuelas públicas son, después de todo, instituciones públicas financiadas con el dinero de los contribuyentes y que la opinión pública tiene un legítimo interés en la calidad de la educación que se imparte en ellas. Es por medio del sistema de evaluación del desempeño docente que los miembros de la opinión pública, a través de sus legisladores, consejos locales de educación y administradores, garantizan la calidad de la enseñanza. En otras palabras, un padre, al confiar la educación de un niño a las escuelas públicas, tiene el derecho a esperar un determinado nivel de desempeño mínimo” (Danielson y McGreal, 2000). Por otra parte, los docentes y sus sindicatos esperan oportunidades de reconocimiento social de su trabajo y oportunidades de desarrollo profesional a través de la implementación de un sistema formativo de evaluación del desempeño docente (Ávalos y Assael, 2006).

**TABLA 1**

**RESUMEN DE LOS ARGUMENTOS EN FAVOR Y EN CONTRA DE LA EVALUACIÓN DOCENTE SEGÚN LO DESCRITO EN LA LITERATURA**

	<b>Argumentos a favor de la evaluación docente</b>	<b>Argumentos en contra de la evaluación docente</b>
<b>Responsabilidad por los resultados</b>	El sistema actual no hace responsables a los docentes por su práctica y su desempeño. Es un ejercicio burocrático sinsentido que requiere cambios.	El diseño de un sistema de evaluación equitativo y riguroso para efectos de responsabilidad por los resultados es vano, debido a que no es posible determinar el desempeño en forma objetiva y la 'buena enseñanza' puede adquirir diversas formas.
	Las autoridades locales y los padres tienen el 'derecho' a instituir mecanismos de garantía de la calidad. La revisión del desempeño aumenta el apoyo político y público a los sistemas educacionales.	La enseñanza requiere un ambiente seguro, alejado de las presiones políticas, sociales y financieras.
	Las evaluaciones permiten la identificación de quienes tienen un buen desempeño de manera similar a otros mercados.	Los mecanismos de mercado no tienen lugar en la educación.
<b>Mecanismos de incentivo y vínculos con el reconocimiento y las retribuciones</b>	Las evaluaciones entregan una base para incrementos de la remuneración que difieren de la escala salarial única basada meramente en la experiencia. Son esenciales para hacer más atractiva la profesión.	Los docentes no están motivados por recompensas financieras, sino por aspectos 'intrínsecos' (por ejemplo, el deseo de enseñar, de trabajar con niños) y condiciones de trabajo favorables (por ejemplo, horarios flexibles). Algunos docentes pueden desmotivarse como consecuencia de los procedimientos de evaluación.
	Existe la necesidad de adoptar las medidas adecuadas ante los docentes ineficientes.	Es una política paternalista que estigmatiza a los docentes.
<b>Desarrollo profesional</b>	Las evaluaciones permiten que los docentes identifiquen sus fortalezas y deficiencias en relación con los objetivos del establecimiento escolar y evalúen sus necesidades de desarrollo profesional. Son esenciales para mantener motivados a los maestros en su trabajo.	La selección de las actividades de desarrollo profesional no debe basarse en los resultados de las evaluaciones, sino que debe ser realizada por el docente unilateralmente. Los establecimientos escolares no ofrecen actividades de desarrollo profesional en las áreas en las cuales se requiere el perfeccionamiento.
<b>Costo</b>	El sistema actual malgasta tiempo, energía y dinero.	Un esquema de evaluación docente integral es oneroso y requiere mucho tiempo.
<b>Efectos</b>	Las evaluaciones mejoran la práctica de los docentes y el aprendizaje de los estudiantes.	Las evaluaciones producen una serie de efectos negativos tales como la restricción del currículo y el descuido de algunos estudiantes.
	Las evaluaciones del desempeño docente mejoran la cooperación entre los docentes –a través de debates profesionales y un intercambio de su práctica– y entre los docentes y los directivos de los establecimientos escolares, de los cuales se espera retroalimentación y <i>coaching</i> .	Las evaluaciones del desempeño docente reducen la cooperación entre los docentes –debido a los efectos de competencia– y entre los docentes y los directivos de los establecimientos escolares, debido a las relaciones jerárquicas o entre 'evaluador-evaluado'.

Por otra parte, los docentes suelen rechazar las reformas diseñadas por las autoridades a cargo de la formulación de las políticas, debido a que las consideran incompatibles con sus propios intereses y prácticas diarias. Kennedy (2005) argumenta que el rechazo a las reformas por parte de los maestros “muy dedicados” no proviene de su falta de disposición al cambio o al mejoramiento, sino del “triste hecho de que la mayor parte de las reformas no reconocen las realidades de la enseñanza en el aula”. Los reformadores tienen altas expectativas con respecto al contenido escolar riguroso y pertinente, al compromiso intelectual y al acceso universal al conocimiento. Sin embargo, las reformas fracasan, principalmente debido a que “las circunstancias de la docencia impiden que los docentes modifiquen sus prácticas”. Por ejemplo, los docentes enfrentan restricciones diarias y hechos inesperados, soportan las presiones del calendario académico y sirven a una ‘clientela obligatoria’ y heterogénea.

En la Tabla 1 se resumen los argumentos en apoyo y en contra de las evaluaciones docentes, de acuerdo a algunas dimensiones claves. Si bien los argumentos en favor de las evaluaciones generalmente provienen de las autoridades a cargo de la formulación de las políticas y los argumentos en contra provienen de los sindicatos docentes, la realidad es menos tajante. Por ejemplo, la Federación Norteamericana de Maestros y la Asociación Nacional de Educación, los dos principales sindicatos docentes norteamericanos (Hesse y West, 2006), promueven la certificación de la NBPTS como “una manera probada de fortalecer las habilidades, conocimientos, profesionalismo y reconocimiento de los maestros” (AFT y NEA, 2008). Sin embargo, como se mencionó anteriormente, la evaluación de la NBPTS es solicitada voluntariamente por los docentes que desean revisar su práctica en función de los más elevados estándares profesionales y no tiene ninguna consecuencia negativa. La respuesta de los sindicatos podría haber sido diferente ante un esquema de evaluación obligatorio generalizado en el cual todos los maestros fuesen evaluados, cualquiera fuese su nivel o calidad, o en el cual pudieran estar en juego potenciales despidos o postergaciones de los ascensos.

Asimismo, se observa una gran heterogeneidad en las respuestas de los docentes a los esquemas de evaluación. Por ejemplo, los docentes

más jóvenes podrían mostrarse más dispuestos a aceptar evaluaciones sumativas y vínculos entre los resultados de las evaluaciones y la remuneración que los maestros más experimentados, que se encuentran en un nivel más alto de la escala salarial. Estudiando la aceptación potencial de la remuneración basada en el desempeño o la remuneración basada en los conocimientos y las habilidades por parte de los nuevos maestros, Milanowski (2007) descubrió que los maestros principiantes podrían visualizar en forma más favorable las nuevas estrategias de remuneración que sus colegas más experimentados. Asimismo, los docentes que han invertido una considerable cantidad de tiempo y dinero en obtener las certificaciones educacionales requeridas por la escala salarial única, exhiben una mayor probabilidad de rechazar la reforma (Odden y Kelley, 2002). Entre los maestros más antiguos, la heterogeneidad también es sorprendente. Day y Gu (2007) descubrieron que los maestros que habían ejercido más de veinticinco años se encontraban en escenarios profesionales extremos: un subgrupo “mostraba un continuo interés por actualizar y mejorar sus conocimientos para el aula”, mientras que el otro subgrupo “mostraba crecientes sentimientos de fatiga y desilusión”. Esta división podría implicar que los primeros estuvieran mucho más dispuestos que los últimos a aceptar –o incluso a promover– una revisión de su práctica y una retroalimentación con respecto a la misma.

### **3.3.2 Cómo superar los obstáculos**

A pesar de los desafíos, es preciso dedicar tiempo y recursos al (re)diseño e implementación de un sistema de evaluación del desempeño docente bien aceptado. De lo contrario, persistirán las preocupaciones asociadas a los actuales procedimientos, desmotivadores y onerosos (Danielson, 2001; Milanowski y Kimball, 2003; Holland, 2005; Marshall, 2005). La literatura destaca la crucial importancia de los siguientes elementos para superar los obstáculos para la implementación.

#### ***Participación en el diálogo y las consultas.***

La concepción inicial del sistema debería incluir la amplia participación de todos los actores clave, especialmente los docentes y sus sindicatos, desde el inicio de las conversaciones (Ávalos y Assael, 2006). Los maestros aceptarán más fácilmente ser evaluados si se los consulta durante el diseño del proceso. Además de tomar sus

temores y demandas en consideración, la participación de los docentes reconoce su profesionalismo, la escasez de sus habilidades y la extensión de sus responsabilidades (Hess y West, 2006), como también su posición indispensable para estimar la factibilidad y relevancia del sistema de evaluación docente. Los docentes son el 'núcleo técnico' de los sistemas educacionales y su compromiso es esencial en el desarrollo de los sistemas de evaluación dada la profundidad de sus conocimientos y prácticas profesionales. Si los procedimientos de evaluación del desempeño docente se diseñan en forma unilateral al nivel de la superestructura administrativa sin considerar e incluir el núcleo de la práctica docente, se producirá una 'conexión suelta' entre los administradores y los docentes, la que no ofrecerá las garantías públicas de calidad y a la vez desincentivará la reflexión y la revisión entre los docentes mismos (Elmore, 2000; Kleinhenz e Ingvarson, 2004). Por lo tanto, los administradores y los sindicatos docentes necesitan trabajar mano a mano en la creación del nivel de confianza y cooperación requerido para permitir el avance de la reforma de manera productiva (Odden y Kelley, 2002).

**Apoyo a los maestros en la comprensión y apropiación de la evaluación.** También es de vital importancia garantizar que los docentes reciban el apoyo necesario para comprender los procedimientos de evaluación. Los maestros deben saber lo que se espera de ellos para ser reconocidos como 'buenos' docentes antes del inicio del proceso. Esto requiere no solo completa transparencia en los criterios y procedimientos de evaluación, sino también garantizar que los maestros se apropien del proceso a través del apoyo y el *coaching*. Por ejemplo, la *Guía para Comprender la Certificación de la Junta Nacional* responde a las preocupaciones de los maestros en relación con las características de la evaluación de la NBPTS. Es importante explicar el sistema (quiénes son los involucrados, en qué consiste el proceso, cómo se establecen los puntajes, etc.) y ofrecer asesoría para ayudar a que los docentes tengan éxito (qué incluir en un portafolio, qué ejercicios preparar, ejemplos e ideas de postulantes y formadores anteriores) (AFT y NEA, 2008).

**Aplicación de una implementación piloto antes de la plena implementación.** La aplicación de una implementación piloto es una forma rentable de garantizar la viabilidad y confiabilidad

del sistema antes de la plena implementación. Asociada a las percepciones obtenidas a partir de los actores clave, permite la revisión del proceso y la introducción de ajustes a la luz de las posibles deficiencias. Heneman *et al.* (2006) sostienen que se requiere al menos un año piloto para resolver los problemas de operación de los sistemas de evaluación. Sin embargo, el traspaso a plena escala después de la ejecución piloto algunas veces revela otros problemas de implementación, los que, a su vez, reducen la credibilidad del sistema ante los docentes y la aceptación por parte de estos. Por lo tanto, es preciso actuar con precaución en la selección de establecimientos o docentes representativos para la implementación piloto.

#### 4. EVIDENCIA EMPÍRICA DE LOS EFECTOS DE LOS ESQUEMAS DE EVALUACIÓN DEL DESEMPEÑO DOCENTE

En esta sección se resume la evidencia empírica existente con respecto a los efectos de la evaluación del desempeño docente en la práctica docente y el aprendizaje de los estudiantes. En primer lugar, se destacan las dificultades enfrentadas para evaluar los efectos cuantitativos asociados a los esquemas de evaluación del desempeño docente y las conclusiones mixtas resultantes. A continuación, se consideran las circunstancias bajo las cuales los sistemas de evaluación del desempeño docente parecen ser más eficaces.

##### 4.1 EVIDENCIA CUANTITATIVA

###### 4.1.1 Dificultades para medir la calidad de los docentes y el impacto de la evaluación docente

Gran parte de la literatura establece que los docentes tienen una alta incidencia en los indicadores de rendimiento escolar, en el sentido de que son poderosos contribuyentes al rendimiento académico de los estudiantes (Vandervoort *et al.*, 2004; OCDE, 2005). Sin embargo, la literatura se muestra más renuente a demostrar cuáles son las características de los maestros que son relevantes para la calidad de los docentes y cuál es la importancia relativa de la calidad de los docentes en comparación con otros factores que teóricamente influyen en el aprendizaje de los estudiantes, incluyendo los factores familiares, los

factores asociados a los estudiantes y los factores escolares. Aaronson, Barrow y Sander (2007) recalcan que la literatura sobre este tema “sigue estando relativamente en la oscuridad” a pesar del mejoramiento de la información. No es de sorprender que la medición del impacto de la evaluación docente en términos del aprendizaje de los estudiantes a través de las ‘funciones de producción educacionales’ sea incluso más difícil.

La medición del efecto de la evaluación docente enfrenta una serie de desafíos. En primer lugar, es preciso controlar el amplio conjunto de variables cualitativas que probablemente influyen en el aprendizaje de los estudiantes. Estas variables comprenden las características del maestro (por ejemplo, edad, género), su formación y experiencia, los factores familiares que afectan al estudiante (por ejemplo, los antecedentes de los padres, el apoyo de los padres), los factores escolares (por ejemplo, las políticas del establecimiento escolar, los incentivos del establecimiento, los efectos de los pares y el aula) y los factores de cada estudiante (por ejemplo, la motivación, las habilidades cognitivas, la experiencia acumulada). Las complejas realidades de la educación impiden que los investigadores asimilen con precisión estos factores como datos de entrada tradicionales en las funciones de producción (Hanushek, 1986). En segundo lugar, debido a su naturaleza cualitativa y heterogénea, el producto en sí mismo –el aprendizaje de los estudiantes– no es un ‘producto final’ tradicionalmente medible y esto hace aún más difícil la descomposición entre los aportes de los diferentes factores (Hanushek, 1986; Ingvarson *et al.*, 2007). Esto no significa que es inútil realizar cualquier estudio cuantitativo en educación, sino que se requiere prestar particular atención a los aspectos analíticos o a las potenciales interpretaciones erróneas de los resultados. Además, es preciso considerar que cada omisión de un factor o problema de medición –incluyendo la ausencia de datos– genera un posible sesgo cuantitativo en la relación estimada entre la calidad de los docentes y el rendimiento académico de los estudiantes (Xin, Xu y Tatsuoka, 2004).

A consecuencia de lo anterior, la literatura empírica, que principalmente indicaba que la evaluación docente podría tener una función importante en el aprendizaje de los estudiantes, derivaba de un proceso de eliminación. Por medio de una oposición o restricción de los respectivos roles de los rasgos evidentes de cada maestro en

particular (ya sean características, formación, experiencia o incentivos financieros), numerosos estudios concluyeron que lo que en realidad importaba eran las prácticas del maestro y, por extrapolación, la evaluación de estas prácticas. El primer aporte influyente fue la distinción de Hanushek entre las características observables de los docentes, tales como los antecedentes, el género o la raza, y las “habilidades” no cuantificables de los docentes (Hanushek, 1986, 1992). De acuerdo a Hanushek, el hecho de que la literatura anterior no haya encontrado ningún impacto significativo de la calidad de los docentes en el rendimiento académico de los estudiantes se ha debido a que se ha concentrado en los atributos observables de los docentes –el hecho de que el docente tuviera un grado de magíster, por ejemplo– mientras que la calidad de los docentes estaba, por el contrario, relacionada con sus “habilidades” o “elecciones idiosincráticas de la pedagogía y los métodos” (tales como el manejo del aula, los métodos de presentación de las ideas abstractas, las habilidades comunicacionales, etc.), es decir, su práctica.

Basándose en el modelo de Hanushek, un cuerpo de nuevos estudios empíricos entrega evidencia de una escasa relación entre los atributos de los docentes y los indicadores de rendimiento escolar para apoyar la evaluación de la práctica docente. Por ejemplo, Muñoz y Chang (2008) destacan la vaga relación entre la formación y los años de experiencia de los docentes y el progreso escolar de los estudiantes, con la intención de elevar el perfil de la revisión de las prácticas pedagógicas de los docentes entre las políticas educacionales. Aaronson, Barrow y Sander (2007) descubrieron que las medidas tradicionales aplicadas a los docentes, incluyendo aquellas que determinan la actual remuneración docente, explican poco de la variación de la calidad estimada. Estos hallazgos cuestionan la actual escala salarial única basada en la formación y la experiencia de los docentes en beneficio de incrementos de la remuneración que consideren determinantes más amplias, incluyendo el vínculo con los sistemas de evaluación de la práctica y el desempeño de los docentes.

#### **4.1.2 Evidencia empírica mixta con respecto a los sistemas de evaluación docente**

A diferencia de los estudios mencionados en la subsección anterior, a continuación se aborda la

evidencia cuantitativa en torno a los sistemas de evaluación docente *per se*. Sin embargo, cabe señalar que la mayor parte de las investigaciones empíricas no se centran en la estimación de los efectos directos sobre los resultados de los estudiantes (ni considerando un sistema de evaluación docente en particular ni en forma longitudinal). Por el contrario, la literatura cuantitativa comprende principalmente las dos categorías siguientes de estudios empíricos.

El primer subgrupo examina la variación en la relación estadística entre los docentes y los resultados de los estudiantes cuando los docentes aprueban un proceso de evaluación particular y cuando no lo hacen. Este cuerpo de evidencia no mide los efectos de la evaluación docente en los resultados de los estudiantes, puesto que compara dos grupos distintos de maestros (uno sujeto a evaluación y el otro no) en lugar de comparar el impacto de un grupo particular de maestros en los resultados de los estudiantes antes y después del proceso de evaluación considerado. Por el contrario, entrega una indicación de la capacidad del proceso de evaluación implementado de diferenciar eficazmente los maestros competentes de aquellos que no lo son. Estos estudios son esenciales, puesto que establecen la viabilidad y confiabilidad de un esquema de evaluación, aspectos indispensables para la equidad de los procedimientos sumativos y el potencial vínculo con las retribuciones.

Por ejemplo, numerosos estudios analizaron la viabilidad del sistema de evaluación de la NBPTS, debido a que representa uno de los enfoques más complejos e integrales con respecto a la evaluación del desempeño docente y se traduce en un reconocimiento formal: la Certificación de la Junta Nacional (*National Board Certification* – NBC). Una serie de autores (Cavalluzzo, 2004; Goldhaber y Anthony, 2007; Vandervoort *et al.*, 2004; Smith *et al.*, 2005) descubrieron que los estudiantes de maestros que han obtenido la NBC exhiben un mejor rendimiento en las pruebas estandarizadas que los estudiantes de maestros no certificados. Esto indica, en primer lugar, que las prácticas docentes son importantes para el rendimiento académico de los estudiantes y, en segundo lugar, que la NBC identifica correctamente a aquellos maestros que han adoptado las mejores prácticas. Más aún, Goldhaber y Anthony (2007) y Cavalluzzo (2004) también concluyen que los puntajes de los estudiantes mejoraron, particularmen-

te en el caso de los estudiantes minoritarios y los estudiantes con necesidades especiales, sugiriendo, por ende, que la NBC identifica adecuadamente a los maestros que adoptan las prácticas que mejoran la equidad educacional además de la eficiencia general. Sin embargo, otros autores (McColskey y Stronge, 2005; Sanders *et al.*, 2005; Harris y Sass, 2007) mostraron que, en contraste, los estudiantes de maestros que habían obtenido la NBC no exhibían un rendimiento significativamente mejor que los demás estudiantes, a pesar de mejoramientos en algunos grados o áreas.

La evidencia empírica también es mixta en el caso de los sistemas de evaluación docente obligatoria. Milanowski (2004) estimó la relación entre las calificaciones derivadas de la evaluación docente y una medida de rendimiento de valor agregado de los estudiantes para el distrito norteamericano de Cincinnati, que ha implementado un esquema de evaluación docente integral basado en estándares como base para un sistema de remuneración basada en los conocimientos y las habilidades. Descubrió significativas correlaciones positivas y concluyó que, si los puntajes derivados de un sistema de evaluación docente riguroso están sustancialmente relacionados con el rendimiento académico de los estudiantes, esto entrega evidencia válida para el uso de los puntajes obtenidos por los maestros como base para un sistema de retribución financiera. Borman y Kimball (2005) estudiaron el sistema de evaluación docente del distrito de Washoe County, que consiste en un modelo de dos niveles. Después de controlar los antecedentes de los estudiantes y la experiencia de los maestros, evaluaron la relación entre la calidad de los docentes medida a través del sistema de evaluación y tanto el rendimiento medio general del curso como los efectos al interior del aula en la equidad social. Descubrieron que los maestros con altos puntajes en las evaluaciones están relacionados con mejores resultados de aprendizaje de los estudiantes en todos los grados y áreas pedagógicas (lectura y matemáticas). Sin embargo, estos maestros no parecen estar reduciendo las brechas en el rendimiento entre los estudiantes con menor y mayor rendimiento y los estudiantes con antecedentes de bajos ingresos o minoritarios. Esta es una fuente de escepticismo al considerar la validez del sistema de evaluación para distinguir entre los maestros que adoptan prácticas orientadas a la equidad y aquellos que no lo hacen.

El segundo subgrupo de estudios cuantitativos con respecto a los sistemas de evaluación docente se centra en los efectos en el mejoramiento de la práctica y la motivación de los maestros, según lo percibido por el maestro evaluado. Si los maestros informan un mejoramiento de las prácticas como resultado del proceso de evaluación y suponiendo que las prácticas correspondientes son relevantes para el aprendizaje de los estudiantes, se supone que el sistema de evaluación es eficaz en mejorar indirectamente los resultados de los estudiantes. Por ejemplo, al solicitar a los maestros la preparación de portafolios y una reflexión acerca de sus prácticas, la NBPTS ofrece una oportunidad rentable de desarrollo profesional a través del proceso de evaluación (Cohen y Rice, 2005), motivando al maestro a centrarse en las características de un currículo sólido y en las características de una evaluación rigurosa del aprendizaje de los estudiantes. Diversos autores (Bond *et al.*, 2000; Lustick y Sykes, 2006) destacan que los maestros aplicaban en el aula lo que habían aprendido a partir del proceso de evaluación. También pareciera que los maestros cobraban un nuevo entusiasmo por la profesión –en relación al tiempo que esperaban permanecer en la docencia– como resultado de haber enfrentado el proceso de evaluación (Vandervoort *et al.*, 2004; Lustick y Sykes, 2006; Sykes *et al.*, 2006; NBPTS, 2007). Finalmente, los maestros competentes que se someten al proceso de evaluación exhiben probabilidades de contribuir al liderazgo escolar mediante la adopción de nuevas funciones, incluyendo tutorías y *coaching* de otros maestros que reconocen la ayuda que pueden recibir de los maestros certificados (Petty, 2002; Freund *et al.*, 2005; Sykes *et al.*, 2006). Estos estudios aportan considerables nociones relativas al aspecto formativo del sistema de evaluación docente.

Como se mencionó anteriormente, existe muy poca evidencia de la correlación directa entre la evaluación del desempeño docente y el rendimiento de los estudiantes. Figlio y Kenny (2007) intentaron introducir una evaluación docente como un mecanismo de incentivo para los maestros orientado al rendimiento académico de los estudiantes –junto con medidas de incentivos financieros– en una regresión longitudinal. Utilizaron la base de datos longitudinal norteamericana de escuelas para distinguir entre las escuelas que evalúan a sus maestros experimentados anualmente de aquellas que lo hacen

con menor frecuencia, con la expectativa de que una revisión más frecuente del desempeño mejorara el desempeño docente. También controlaron un amplio conjunto de variables de los estudiantes y de las escuelas, además de los incentivos financieros para los maestros y la amenaza de despidos. Lamentablemente, volvieron a encontrar que la evaluación del desempeño docente no era estadísticamente significativa, mientras que los incentivos financieros estaban correlacionados positiva y significativamente con el rendimiento de los estudiantes.

Los resultados de Figlio y Kenny destacan las dificultades para estimar el impacto directo de la evaluación del desempeño docente en el rendimiento de los estudiantes. Sin embargo, el hecho de que las retribuciones financieras se identifiquen como mecanismos de incentivos positivos orientados al éxito de los estudiantes sugiere que un sistema de evaluación del desempeño docente es indispensable, puesto que permite la identificación de los maestros eficientes que recibirán una retribución. Además, puede que no sea la frecuencia de evaluación, sino la calidad y la sofisticación del proceso de evaluación lo que importe. Esto explica probablemente la razón por la cual los estudios empíricos suelen centrarse en el diseño de esquemas de evaluación exitosos en lugar de mediciones cuantitativas de sus impactos.

## 4.2 EVIDENCIA CUALITATIVA

La evidencia también indica que la evaluación docente parecer ser más eficaz bajo ciertas circunstancias particulares.

***Participación de los docentes en cada etapa del proceso.*** Es preciso consultar a los docentes con respecto a las fortalezas y deficiencias del sistema desde su diseño hasta su plena implementación y revisión. En primer lugar, los docentes deben estar de acuerdo con el marco que define los estándares de la profesión. “La participación activa de los docentes en la construcción y perfeccionamiento del modelo es esencial” (Heneman *et al.*, 2006). El *Marco para la Enseñanza* de Danielson (1996, 2007) o los estándares desarrollados por la NBPTS podrían utilizarse como puntos de partida, con ajustes posteriores para cumplir con los objetivos educacionales locales. La creación de Consejos de Docencia, como en Irlanda en el período 2005-2006, ofrece grandes oportunidades para involu-

crar a los docentes en el establecimiento de estándares profesionales de alto nivel y, de manera más general, para integrar plenamente a los maestros en la (re)definición de la profesión para el posterior desarrollo de las políticas (OCDE, 2005).

En segundo lugar, *todos* los docentes deben recibir apoyo para comprender lo que la evaluación espera de ellos para reconocerlos como buenos maestros y para prepararse adecuadamente para el proceso de evaluación. Esto requiere tanto transparencia con respecto a los métodos utilizados como también *coaching* orientado a una evaluación con empoderamiento. En palabras de Heneman *et al.* (2006): “En el caso de los maestros, la formación inicial debería centrarse en la naturaleza de las competencias de desempeño en las cuales se basa el sistema, los propósitos y mecanismos del sistema de evaluación y los conocimientos y habilidades necesarios para operar eficazmente bajo el nuevo sistema”. Podrían integrarse expectativas claras con respecto al proceso de evaluación y el correspondiente perfeccionamiento docente en las acciones de liderazgo de los directores de las escuelas (Ovando y Ramírez, 2007). La necesidad de eliminar la ‘conexión suelta’ entre la superestructura administrativa y el núcleo técnico conformado por los docentes (Ingvarson *et al.*, 2007), es decir, la transición desde los procedimientos burocráticos actuales bajo los cuales los docentes son *sometidos* a una evaluación hasta una evaluación realizada *con* los docentes, requiere la plena apropiación de la reforma por parte de los docentes.

Finalmente, también es preciso ofrecer a los docentes oportunidades de expresar sus percepciones y preocupaciones con respecto al proceso de evaluación después de la instauración del sistema. Las entrevistas y encuestas son métodos comunes utilizados por los analistas para recoger las reacciones de los docentes con respecto al sistema de evaluación. Las preguntas generalmente incluyen la comprensión del proceso, la aceptación de los estándares, la equidad del proceso y de los resultados, la capacidad y objetividad del evaluador, la calidad de la retroalimentación recibida, el impacto percibido del proceso de evaluación sobre la enseñanza y la impresión general del sistema (Milanowski y Heneman, 2001, 2004; Kimball, 2002). Milanowski y Heneman (2001) descubrieron que la actitud general favorable de los docentes hacia

un sistema recientemente implementado en un distrito escolar norteamericano de mediano tamaño estaba correlacionada con la aceptación de los estándares docentes, la equidad percibida del proceso, las cualidades del evaluador y la percepción de que el sistema de evaluación tiene un impacto positivo en su enseñanza. Los docentes recibieron 1,5 días de perfeccionamiento para comprender los dominios y estándares evaluados y familiarizarse con los aspectos que serían considerados por los evaluadores y se les proporcionó información con respecto al desarrollo de un portafolio.

**Formación de los evaluadores.** La literatura apunta a un amplio consenso con respecto a la necesidad de una exhaustiva formación de los evaluadores. En primer lugar, los evaluadores deben recibir capacitación para calificar a los docentes de acuerdo a la evidencia limitada que recopilan, los criterios de una buena enseñanza y los correspondientes niveles de calidad de los docentes. Esto es particularmente importante cuando los evaluadores son los directores de los establecimientos escolares, que podrían tener conocimientos limitados con respecto a los contenidos y las habilidades pedagógicas requeridas para la materia enseñada por el docente evaluado. En segundo lugar, debe ofrecerse capacitación a los evaluadores en la entrega de retroalimentación constructiva y *coaching* a los docentes para el futuro mejoramiento de la práctica.

**Liberación de otras tareas, tanto para los evaluadores como para los docentes.** Los sistemas de evaluación docente integrales requieren tiempo y otros recursos. Esto puede ser oneroso, pero es indispensable para el diseño de un sistema coherente y equitativo aprobado y aceptado como propio por los docentes. Una consecuencia es que tanto los docentes como los evaluadores deberían ser liberados parcialmente de otras labores. Milanowski y Heneman (2001) descubrieron que incluso si los docentes aceptan los estándares y la necesidad de un sistema de evaluación, podrían manifestar reticencia cuando el sistema recarga demasiado su carga de trabajo. Por lo tanto, los docentes deberían contar con el tiempo para reflexionar con respecto a su práctica, especialmente cuando el proceso exige la preparación de un portafolio. Como lo enfatiza Heneman *et al.* (2006): “Los diseñadores del sistema deben analizar cuidadosamente qué se requiere de los maestros para minimizar la carga. (...) Tal vez alguna pe-

queña reducción de las demás responsabilidades durante el período en el cual los maestros están siendo sometidos a la evaluación disminuiría la percepción de carga y sensación de estrés”. Las autoridades a cargo de la formulación de las políticas también deberían intentar reducir la carga de trabajo administrativa de los evaluadores, especialmente de los directores de los establecimientos escolares, con el fin de otorgarles más tiempo para la evaluación de los docentes, la retroalimentación y el *coaching* (Marshall, 2005).

***Ejecución de una implementación piloto y revisión permanente del proceso.*** Una implementación piloto constituye una forma rentable de garantizar que el sistema sea eficiente, equitativo y coherente con las necesidades locales antes de la plena implementación. Las entrevistas a los maestros durante la implementación piloto son esenciales para corregir las posibles deficiencias e inquietudes con respecto al sistema. Los investigadores o profesionales también deberían concentrarse en estudios de validez y confiabilidad. Pecheone y Chung (2006) muestran que la consistencia de los puntajes debe ser minuciosamente examinada. Por ejemplo, los puntajes basados en estándares son relevantes si corresponden a las calificaciones holísticas del desempeño de los docentes y si supervisores familiarizados con los docentes muestran un amplio consenso con respecto al nivel de desempeño. Milanowski y Heneman (2001) y Milanowski (2004) recalcan la necesidad de controlar la confiabilidad de los evaluadores mediante un ‘proceso de calibración’ consistente en comparar las calificaciones experimentales de algunos evaluadores con jueces expertos y analizando las diferencias. Solo debe permitirse que evaluadores acreditados evalúen a los maestros después de la plena implementación. También deberían realizarse revisiones del proceso después de la plena implementación. Es más probable que los maestros acepten el proceso en forma rápida si saben que tendrán la oportunidad de expresar sus preocupaciones y opiniones con respecto a las adecuaciones necesarias a medida que el proceso evoluciona.

***Incorporación de la evaluación del desempeño docente en las políticas de calidad y apoyo docente más generales.*** La evaluación no debería reemplazar otros medios de garantizar la calidad de los docentes, tales como la formación docente y los programas de licenciatura,

los programas de inducción, el desarrollo profesional, la retroalimentación y asesoría informales permanentes y el reconocimiento más amplio de la experticia y el compromiso con el trabajo exhibidos por los docentes (AFT, 2001; Corcoran, 2007).

## 5. CONCLUSIÓN

En la primera sección de este informe se da énfasis a la amplia diversidad de esquemas de evaluación del desempeño docente existente en los países de la OCDE. Los países exhiben grandes diferencias en cada una de las características relevantes, incluyendo los respectivos roles de los actores clave, el alcance y propósito de la evaluación, los métodos e instrumentos utilizados para evaluar a los docentes como también los criterios de una ‘buena’ enseñanza y los vínculos con las retribuciones o el desarrollo profesional. Además de los debates acerca de las características particulares y las consecuencias para las carreras profesionales de los maestros, la evaluación del desempeño docente parece ser particularmente conflictiva en los países en los cuales la evaluación docente deriva de procedimientos burocráticos impuestos en lugar de constituir una parte integral de las políticas docentes y escolares más generales.

Cada vez son más los países que muestran un creciente interés por implementar sistemas integrales de evaluación del desempeño docente en respuesta a las demandas de una alta calidad educacional. A pesar de que actualmente se dispone de poca evidencia empírica, la literatura concuerda principalmente en la necesidad de aclarar el propósito enfatizado y la importancia de incluir un conjunto diverso de evaluadores y criterios para reflejar de mejor manera la complejidad de la definición de lo que es una buena enseñanza. También existe un amplio consenso con respecto a la participación de los maestros durante todo el desarrollo del proceso de evaluación. Un esquema de evaluación eficiente, equitativo y confiable requiere la aceptación general y apropiación del sistema por parte de los docentes. El desarrollo de un enfoque integral podría ser oneroso, pero es crítico para conciliar las demandas de calidad educacional, el perfeccionamiento de las prácticas docentes mediante el desarrollo profesional y el reconocimiento de los conocimientos, habilidades y competencias de los docentes.

## ANEXO 1

### MARCO CONCEPTUAL PARA LA EVALUACIÓN DEL DESEMPEÑO DOCENTE

#### Principales organismos u organizaciones involucrados / Actores:

- |   |  |
|---|--|
| <ul style="list-style-type: none"> <li>- Gobiernos nacionales (Ministerios / Departamentos de Educación)</li> <li>- Autoridades descentralizadas a cargo de las políticas educacionales (distritos, municipalidades)</li> </ul> | <ul style="list-style-type: none"> <li>- Directivos escolares</li> <li>- Maestros y sindicatos docentes</li> <li>- Padres / Estudiantes</li> </ul> |
|---|--|

#### Alcance de la evaluación / Maestros evaluados:

- Todo el país *versus* procedimientos regionales
- Tipo de establecimiento: escuelas públicas, escuelas privadas
- Periodicidad de la evaluación: parte del trabajo regular *versus* evaluación en casos especiales (ascensos, reclamos)
- Obligatoria *versus* voluntaria
- Todos los docentes están sujetos a la misma evaluación *versus* evaluación adaptada de acuerdo a la experiencia del docente
- Implementación piloto *versus* plena implementación

#### Evaluadores:

- Revisiones internas (por los directores o el personal directivo del establecimiento)
- Revisiones externas (por parte de pares o maestros competentes vinculados a la misma área de contenidos pedagógicos)
- Autoevaluación
- Padres
- Estudiantes



#### Criterios y estándares:

- Conocimiento de los contenidos en la materia enseñada.
- Habilidades pedagógicas
- Conocimiento de los estudiantes
- Capacidad de mejorar el rendimiento de los estudiantes
- Competencia en la planificación pedagógica
- Conocimientos sobre medición del aprendizaje de los estudiantes
- Habilidad para crear un ambiente favorable en el aula
- Habilidad para manejar los procedimientos de aula
- Capacidad para involucrar a los estudiantes en el aprendizaje e interactuar con ellos
- Habilidades de comunicación y supervisión
- Habilidades para satisfacer las necesidades de poblaciones de estudiantes diversificadas; demostración de flexibilidad y sensibilidad
- Profesionalismo: comunicación con las familias, el personal docente del establecimiento y los directivos
- Participación en actividades de perfeccionamiento y desarrollo profesional: reflexión sobre la enseñanza, formación continua



**EVALUACIÓN  
DE LA  
PRÁCTICA  
Y EL  
DESEMPEÑO  
DE DOCENTES**

#### Métodos e instrumentos:

- Observaciones en el aula
- Entrevistas con el docente
- Portafolios preparados por los docentes (videoclips, planes de lecciones, hojas de reflexión, cuestionarios autoadministrados, muestras de trabajos de los estudiantes)
- Resultados en términos de rendimiento académico de los estudiantes (desempeño absoluto o mejoramientos de valor agregado)
- Pruebas a los docentes
- Datos de cuestionarios y encuestas respondidas por los padres y los estudiantes



#### Evaluación sumativa:

- **Responsabilidad por los resultados y garantía de calidad para las autoridades a cargo de las políticas y los padres**
  - Mejoramiento del aprendizaje y el rendimiento de los estudiantes a través de mejores prácticas docentes
  - Reducción de la inequidad en el rendimiento de los estudiantes
- **Reconocimiento y/o retribuciones para los docentes:**
  - Reconocimiento de las habilidades y el compromiso
  - Ascenso
  - Incrementos salariales
  - Retribuciones no financieras (condiciones laborales)
  - Medidas ante los docentes ineficientes (postergaciones del ascenso, despidos)
- **Convertir a la docencia en una elección de carrera atractiva**
- **Retención de los maestros eficientes en las escuelas**

#### Evaluación formativa:

- **Desarrollo profesional para mejorar la enseñanza:**
  - Identificación de las fortalezas y deficiencias del docente
  - Entrega de retroalimentación constructiva sobre las prácticas del docente
  - Asesorías a los docentes con respecto a los programas de desarrollo profesional adecuados y las oportunidades de perfeccionamiento de sus capacidades
- **Mantenimiento de la motivación de los docentes durante toda su carrera**
- **Mejoramiento del liderazgo escolar:**
  - Adaptación de los programas de desarrollo profesional de los establecimientos escolares a las necesidades identificadas
  - Mejoramiento de la supervisión y *coaching* de los docentes por parte de los directores
- **Participación de los docentes en el desarrollo e implementación de las políticas**

## ANEXO 2

### EJEMPLOS DE SISTEMAS DE EVALUACIÓN DEL DESEMPEÑO DOCENTE EN LOS PAÍSES DE LA OCDE

#### 1. EVALUACIÓN DOCENTE CON PROPÓSITOS SUMATIVOS Y VÍNCULOS CON LA REMUNERACIÓN: EL DISTRITO NORTEAMERICANO DE CINCINNATI [MILANOWSKI, 2004]

**Contexto.** Cincinnati es un gran distrito urbano con 48.000 estudiantes y 3.000 maestros en más de 70 establecimientos escolares y programas. Su nivel promedio de rendimiento académico de los estudiantes es bajo en comparación con los distritos suburbanos circundantes. Cincinnati también ha tenido una historia de actividades de reforma escolar, incluyendo la introducción de nuevos diseños a nivel de todo el establecimiento, el diseño de presupuestos basados en las escuelas y la formación de equipos a cargo de la administración de los establecimientos escolares y de impartir la instrucción. La relación entre el sindicato docente y la administración ha sido en general positiva. Como muchos otros distritos urbanos, los programas estatales de responsabilidad por los resultados y las expectativas de la opinión pública han ejercido presión en el distrito para que mejore los indicadores de rendimiento escolar.

**Implementación.** En respuesta a la obsolescencia del sistema de evaluación del desempeño docente existente y a ambiciosos objetivos de mejoramiento del rendimiento académico de los estudiantes, el distrito diseñó un sistema de remuneración basada en los conocimientos y las habilidades y un nuevo sistema de evaluación del desempeño docente durante el año escolar 1998-1999. El sistema de evaluación fue implementado en forma piloto en el año escolar 1999-2000 y se utiliza para la evaluación docente a nivel de todo el distrito desde el año escolar 2000-2001.

**Criterios.** El sistema de evaluación se basa en un conjunto de estándares docentes derivados del *Marco para la Enseñanza* (Danielson, 1996). Diecisiete estándares de desempeño están agrupados en cuatro dominios: (i) planificación y preparación; (ii) generación de un ambiente conducente al aprendizaje; (iii) enseñanza orientada al aprendizaje; y (iv) profesionalismo. Para cada estándar, un conjunto de escalas de calificación de tipo conductista, llamadas rúbricas, describen cuatro niveles de desempeño: insatisfactorio, básico, competente y destacado.

**Instrumentos.** Los maestros son evaluados utilizando las rúbricas, de acuerdo a dos fuentes de evidencia principales: seis observaciones en el aula y un portafolio preparado por el maestro. El portafolio incluye componentes tales como planes de lecciones y unidades, registros de asistencia, trabajos de los estudiantes, registros de contacto con las familias y documentación de las actividades de desarrollo profesional.

**Evaluadores.** Un evaluador docente contratado entre las filas del personal docente y liberado de impartir docencia en el aula durante tres años realiza cuatro observaciones en el aula. Los directores y subdirectores realizan otras dos observaciones.

**Totalización de los puntajes.** Sobre la base de resúmenes de las seis observaciones, los evaluadores docentes realizan una calificación sumativa final en cada uno de los estándares incluidos en los dominios (ii) y (iii), mientras que los directores y los subdirectores califican a los maestros en los estándares incluidos en los dominios (i) y (v), principalmente basándose en el portafolio del maestro. A continuación, se totalizan las calificaciones para cada uno de los

estándares para obtener un puntaje para cada uno de los cuatro dominios.

**Alcance y frecuencia de la evaluación.** Todo el sistema de evaluación se utiliza para una evaluación integral de los docentes en su primer y tercer año y cada cinco años de ahí en adelante. Se realiza una evaluación menos intensiva en todos los demás años, la que es llevada a cabo por los directores y los subdirectores y se basa en evidencia más limitada. La intención de la evaluación anual es constituir tanto una oportunidad para el desarrollo profesional docente como una evaluación para efectos de responsabilidad por los resultados.

**Capacitación en el proceso de evaluación.** Tanto los docentes como los evaluadores reciben considerable capacitación en el nuevo sistema. Los evaluadores son capacitados utilizando un proceso de calibración que comprende lecciones grabadas sobre calificación aplicando las rúbricas y comparando luego las calificaciones con jueces expertos y analizando las diferencias. Para garantizar la consistencia entre los evaluadores, el distrito finalmente requiere que todos los evaluadores, incluidos los directores, cumplan un determinado nivel de concordancia con un conjunto de referencias o evaluadores expertos en lecciones de calificación grabadas en video. A partir del año escolar 2001-2002, solo quienes cumplen los estándares están autorizados para evaluar.

**Consecuencias directas.** Para los maestros principiantes (aquellos evaluados durante su primer y tercer año), la consecuencia de una baja evaluación global podría ser el término del contrato. En el caso de los docentes permanentes, las consecuencias de una evaluación positiva podrían incluir la elegibilidad para convertirse en un líder pedagógico. Una mala evaluación podría llevar a la asignación a un programa de asistencia entre pares y finalmente a un término del contrato.

**Vínculo con la remuneración.** El sistema de evaluación del desempeño se diseñó en parte para proporcionar la base de un sistema de remuneraciones basado en los conocimientos y las habilidades. Este sistema define los niveles de la carrera profesional docente con una remuneración diferenciada por nivel. El nuevo sistema de remuneraciones se programó originalmente para entrar en vigencia durante el año escolar 2002-2003, traducéndose en evaluaciones de 'altas consecuencias' para los maestros del distrito. Sin embargo, el vínculo entre el sistema de evaluación y la remuneración fue rechazado por los maestros en una elección especial realizada en mayo de 2002.

#### 2. EVALUACIÓN DOCENTE CON PROPÓSITOS FORMATIVOS Y COMO PARTE DE LAS POLÍTICAS ESCOLARES MÁS GENERALES

##### 2A. FINLANDIA [UNESCO, 2007]

**Contexto.** En Finlandia, los maestros de la educación primaria y secundaria tienen cargos comparables a los de los funcionarios públicos nacionales o municipales. Sin embargo, los directivos escolares están a cargo de la selección de los maestros –una vez que han obtenido la licencia requerida– y a cargo de todas las políticas que son consideradas necesarias para el mejoramiento de la calidad de la enseñanza, entre las cuales se encuentra la evaluación del desempeño docente. Finlandia es un caso paradigmático en el

cual el sistema anterior de 'inspección y supervisión de los docentes y los establecimientos escolares' fue eliminado en 1990, pero no fue reemplazado por ningún otro sistema externo similar. Como consecuencia, la evaluación docente actualmente corre en forma paralela con otras políticas en cada escuela particular.

**Métodos/Evaluadores.** El esquema finlandés de evaluación docente se caracteriza por el muy alto nivel de confianza asignado a la escuela y a las competencias y el profesionalismo de los docentes como base para el mejoramiento de la calidad de la enseñanza. Por lo tanto, la autoevaluación de los maestros es considerada uno de los principales medios de optimización profesional. Los directivos escolares también desempeñan una función crucial en comprometer a los maestros en la auto-reflexión acerca de su propia práctica y en desarrollar una cultura de evaluación junto con ambiciosos objetivos, de acuerdo al contexto y los problemas de cada escuela. La mayoría de las escuelas han implementado debates anuales entre los directivos escolares y los docentes para evaluar el cumplimiento de los objetivos personales establecidos durante el año anterior y para establecer nuevos objetivos personales que sean concordantes tanto con el análisis del maestro como con las necesidades de la escuela.

## 2B. INGLATERRA [OFSTED, 2006; TDA, 2007]

**Contexto.** El sistema inglés se diseñó originalmente con propósitos sumativos, teniendo como objetivo la evaluación del desempeño de los docentes y la entrega a estos de oportunidades para acceder a una etapa superior de la carrera profesional y a la escala de remuneraciones correspondiente. Sin embargo, se plantearon numerosas inquietudes con respecto a la equidad del proceso y a los potenciales impactos perversos del procedimiento en el desempeño de los docentes (Kleinhenz e Ingvarson, 2004). Por lo tanto, los recientes avances del sistema, incluyendo nuevos estándares profesionales a partir de septiembre de 2007, indican un mayor enfoque formativo, sustentado por una disposición a reforzar el vínculo entre el sistema de evaluación docente y las necesidades de desarrollo profesional de los docentes en relación con los objetivos de las escuelas. De manera más general, el sistema, implementado dentro de un marco más amplio para toda la fuerza laboral de los establecimientos escolares, tiene por objetivo mejorar el liderazgo escolar y ser una parte integral de las políticas más generales de las escuelas.

**Alcance/Métodos.** La evaluación está diferenciada de acuerdo a la etapa de la carrera profesional del docente evaluado. Se identifican cinco etapas profesionales: (i) la certificación como Maestro Calificado (Q); (ii) los maestros incluidos en el escalafón principal (C); (iii) los maestros incluidos en el tramo superior de la escala de remuneraciones (Maestros Post-Umbra) (P); (iv) los Maestros Excelentes (E); y (v) los Maestros con Habilidades Avanzadas (A).

**Criterios.** En cada etapa, los estándares profesionales docentes comprenden tres dominios. El primero se refiere a los *atributos profesionales* del docente, incluyendo las relaciones con los niños y los jóvenes, la actitud frente al marco y la implementación de nuevas políticas escolares, la comunicación y trabajo con los demás y las actividades de desarrollo profesional. El segundo dominio está compuesto por los *conocimientos profesionales y la comprensión de la profesión* por parte del docente, incluyendo los conocimientos de pedagogía y aprendizaje; la comprensión de la evaluación y la supervisión; las materias y el conocimiento del currículo; las habilidades de alfabetización, habilidades básicas de matemáticas y habilidades en tecnología de la información y la comunicación; la comprensión de los factores que afectan el rendimiento de grupos de estudiantes diversi-

ficados; y los conocimientos sobre salud y bienestar de los estudiantes. El último dominio se refiere a las *habilidades profesionales* de los docentes, incluyendo la planificación, la enseñanza, la evaluación, la supervisión, la entrega de competencias de retroalimentación; la capacidad de revisar y adaptar la enseñanza y el aprendizaje; la capacidad para crear un ambiente de aprendizaje; las capacidades para desarrollar trabajo en equipo y colaboración. Todos estos estándares son representaciones de una buena enseñanza que no reemplazan los deberes y las responsabilidades profesionales de los docentes.

**Consecuencias para el desarrollo profesional de los docentes y vínculos con las expectativas y políticas escolares.** Los estándares apoyan a los docentes en la identificación de sus necesidades de desarrollo profesional. Cuando los docentes desean avanzar a la siguiente etapa de la carrera profesional, el nivel siguiente del marco ofrece un punto de referencia para todos los docentes al considerar su desarrollo futuro. Mientras que no todos los docentes desean necesariamente pasar a la siguiente etapa de la carrera profesional, los estándares también los apoyan en la identificación de las maneras de ampliar y profundizar su experticia en su actual etapa de la carrera profesional. Estos marcos constituyen una base para la responsabilidad profesional y la obligación contractual de comprometer a todos los docentes en un desarrollo profesional eficaz, sostenido y relevante durante toda su carrera profesional. Ofrecen un conjunto continuo de expectativas acerca del nivel de compromiso con el desarrollo profesional, que ofrece claridad y la diferenciación apropiada para cada etapa de la carrera profesional. También establecen expectativas con respecto al aporte que cada uno de los docentes realiza a los demás, tomando en cuenta sus niveles de habilidades, experticia y experiencia y su función en la escuela y reflejando su uso de conocimientos de las materias y pedagogías actualizadas. En todos estos casos, el manejo del desempeño es el proceso clave que ofrece el contexto para debates regulares con respecto a las aspiraciones que tienen los docentes con respecto a su carrera profesional y su desarrollo futuro, en su actual etapa de carrera o después de ella.

Para mayor información:

- Training and Development Agency for Schools (TDA): <http://www.tda.gov.uk/teachers/professionalstandards.aspx> y <http://www.tda.gov.uk/teachers/continuingprofessionaldevelopment.aspx>
- Office for Standards in Education (Ofsted): <http://www.ofsted.gov.uk>

## 3. CONCILIACIÓN DE LOS PROPÓSITOS SUMATIVOS Y FORMATIVOS EN UN ENFOQUE INTEGRAL: CHILE [ÁVALOS Y ASSAEL, 2006]

**Contexto.** El contexto histórico del sistema educacional chileno sin duda ha desempeñado una función crítica en la comprensión de la necesidad de un esquema de evaluación docente integral y conciliador. En 1980, el gobierno militar [1973-1990] transfirió la gestión de las escuelas a las autoridades municipales, lo que implicó un cambio en la condición de los maestros, que pasaron de ser funcionarios públicos a empleados asalariados de las municipalidades. Al término del régimen dictatorial, una de las principales preocupaciones fue el hecho de que las condiciones de los docentes no evolucionaron en forma paralela a aquellas de los funcionarios públicos, lo que tuvo un enorme impacto en la percepción y valoración de la docencia tanto por parte de los docentes como de la opinión pública. En la década de los noventa, la profesión docente sufrió un drástico deterioro de la calidad de los postulantes a la carrera y un empeoramiento

to de las condiciones laborales. Al mismo tiempo, la evidencia de resultados insatisfactorios en términos de aprendizaje de los estudiantes ejerció una fuerte presión sobre el gobierno para que incluyera una cláusula en el nuevo Estatuto Docente (1991) que exigió una evaluación anual de los maestros. Sin embargo, si bien los maestros siguieron solicitando mejores sueldos y condiciones laborales, rechazaron la implementación del sistema de evaluación. Esto fue seguido por un largo período de debates y negociaciones con respecto al modelo de evaluación docente que debería implementarse.

**Diseño e implementación del sistema.** El sistema fue promulgado por ley en agosto de 2004, es decir, alrededor de siete años después del inicio de los debates. El sistema está orientado al mejoramiento de la enseñanza y los resultados de aprendizaje. Su objetivo es estimular a los maestros a continuar su propio perfeccionamiento a través del conocimiento de sus fortalezas y deficiencias. Se basa en criterios explícitos de lo que se evalúa, pero sin prescribir un modelo de enseñanza. Está fundamentado en la articulación de sus diferentes elementos: criterios sancionados por la fuerza laboral docente, una estructura de gestión independiente, evaluadores especialmente capacitados y un conjunto coordinado de procedimientos para recopilar la evidencia requerida por los criterios.

**Principales actores del sistema.** El Centro de Perfeccionamiento, Experimentación e Investigación Pedagógica del Ministerio de Educación está a cargo de la gestión del sistema. Un comité consultivo compuesto por académicos y representantes del Sindicato Docente, la Asociación Chilena de Municipalidades y el Ministerio de Educación supervisa y brinda asesoría con respecto al proceso. Se ha contratado un centro universitario para implementar el proceso: producción y revisión de los instrumentos, selección y preparación de los evaluadores y los profesionales encargados del cálculo de los puntajes y análisis de la evidencia recopilada en cada proceso de evaluación. El proceso de postulación en sí mismo está descentralizado, de manera que, en cada municipalidad, existe un comité que es directamente responsable de organizar los procedimientos de evaluación. La evidencia recopilada se procesa a nivel de las municipalidades y se envía a la unidad de procesamiento central en la universidad, junto con la información contextual que puede ayudar a interpretar los resultados. Esta forma centralizada de procesamiento de la evidencia se implementó a solicitud de los maestros con el propósito de lograr una mayor objetividad.

**Criterios.** El Ministerio de Educación fue la principal instancia en la definición de los criterios de evaluación, obteniendo un conjunto de estándares basados en el trabajo realizado anteriormente con respecto a los estándares de formación docente inicial y el *Marco para la Enseñanza* de Danielson. El resultado es un marco para la enseñanza competente, formulado a través de cuatro dominios de la enseñanza (planificación, ambiente de aprendizaje, profesionalismo y estrategias pedagógicas para el aprendizaje de todos los estudiantes) y veinte criterios/estándares. El marco fue objeto de amplias consultas entre los docentes hasta que se llegó a un acuerdo. Los criterios están vinculados a cuatro niveles de calidad/desempeño: 'insatisfactorio', 'básico', 'competente' y 'excelente'.

**Instrumentos.** La evidencia utilizada para evaluar a los docentes, estructurada en torno al Marco, incluye cuatro fuentes: (i) un portafolio con muestras del trabajo de los maestros y un video de una de sus lecciones; (ii) un formulario de autoevaluación estructurado; (iii) una entrevista estructurada con un evaluador par; y (iv) un informe de las autoridades administrativas y pedagógicas del establecimiento escolar. La evaluación se realiza cada cuatro años.

**Capacitación de los evaluadores.** Los evaluadores pares reciben capacitación específica para su tarea y deben aprobar una prueba para obtener la acreditación. Si bien deberían estar familiarizados con el contexto en el cual se desempeña el maestro evaluado (es decir, condiciones socioeconómicas y laborales), no deben enseñar en el mismo establecimiento escolar.

**Consecuencias de la evaluación.** Uno de los principales problemas que debieron abordarse durante el proceso de negociación se refería a las posibles consecuencias para el docente evaluado. Se acordó que los docentes calificados en un nivel 'básico' recibirían oportunidades específicas de desarrollo profesional con el fin de perfeccionarse. Los docentes calificados como maestros con un desempeño 'insatisfactorio' también reciben oportunidades de desarrollo profesional, pero se los evalúa nuevamente al año siguiente; si el maestro no logra desempeñarse satisfactoriamente en dos evaluaciones consecutivas, es despedido. En contraste, los maestros evaluados como 'competentes' o 'excepcionalmente competentes' tienen prioridad en las oportunidades de ascenso y en las actividades de desarrollo profesional de su interés. También pueden postular a bonificaciones salariales siempre y cuando den una prueba sobre conocimientos curriculares y pedagógicos. El sistema tiene elementos tanto sumativos como formativos en lugar de estar principalmente orientado a uno de los propósitos, lo cual es un resultado del proceso de negociación, el que tomó en consideración los múltiples intereses de los actores. Por ejemplo, los elementos sumativos no incluyen un vínculo entre el desempeño del maestro y los resultados de los estudiantes (un aspecto al cual el sindicato se opuso firmemente) como tampoco un vínculo con el escalafón profesional. El vínculo con el desarrollo profesional es enfatizado y diferenciado sobre la base del nivel de desempeño del maestro.

Para mayor información: Leyes chilenas sobre el Estatuto Docente: Ley N° 3.500; Ley N° 19.070; Ley N° 19.933; Ley N° 19.961.

#### 4. EVALUACIÓN DOCENTE DERIVADA DE PROCEDIMIENTOS BUROCRÁTICOS: FRANCIA [HAUT CONSEIL DE L'ÉVALUATION DE L'ÉCOLE, 2003; PO-CHARD, 2008]

**Contexto.** Los maestros franceses están clasificados en tres categorías distintas de acuerdo a su formación y su certificación inicial: maestros de educación primaria (*professeurs des écoles*), maestros de educación secundaria con certificación regular (*enseignants certifiés*) y maestros de educación secundaria con un mayor nivel de certificación (*enseignants agrégés*). Todos los maestros son funcionarios públicos, pero están ubicados en una de las tres trayectorias profesionales anteriores. Estas difieren en términos de las condiciones y horario de trabajo, la escala de remuneración administrativa y la práctica docente (maestros de educación primaria para múltiples materias *versus* maestros de educación secundaria especializados en una materia). En general, Francia no sufre de escasez de maestros y los exámenes requeridos para ingresar a la profesión siguen siendo selectivos. Sin embargo, Francia tiene problemas en relación con el prestigio social de la docencia y las habilidades necesarias para responder a las necesidades de las escuelas. El actual sistema de evaluación del desempeño docente suele ser descrito como 'no muy justo', 'no muy eficiente' y 'generador de malestar y algunas veces sufrimiento' tanto para los maestros evaluados como para los evaluadores, debido a que está basado en procedimientos administrativos más que en un esquema integral con un claro propósito de mejoramiento.

**Periodicidad de la evaluación/evaluadores.** Se supone que la evaluación del desempeño docente se realizará en

forma regular, como una parte integral del trabajo y los deberes del maestro. Los maestros de educación primaria son evaluados por un inspector docente (*inspecteur*), mientras que los maestros de nivel secundario son evaluados por un panel compuesto por un inspector, que define el 60% del puntaje final, y el director de la escuela, responsable del 40% restante. Sin embargo, la pretendida frecuencia de las evaluaciones suele no cumplir las expectativas. En primer lugar, la frecuencia de las evaluaciones no está establecida legalmente y es determinada arbitrariamente por la disponibilidad de inspectores. Esta es una causa de preocupación en lo que respecta a la equidad del sistema –debido a que maestros que trabajan bajo las mismas condiciones reciben retroalimentación a diversos intervalos– como también en relación con su eficiencia: el intervalo promedio entre dos evaluaciones corresponde a 3-4 años para la educación primaria y 6-7 años para la educación secundaria, un tiempo que se estima demasiado prolongado. Además, la carga de trabajo es tal que podrían surgir preocupaciones con respecto al valor de la retroalimentación. Un inspector se responsabiliza por 350 a 400 maestros, un número excesivo para que la retroalimentación sea eficaz en el mejoramiento de las prácticas de los docentes. En consecuencia, los inspectores mismos informan malestar y frustración asociados al proceso de evaluación, principalmente debido a que sienten que tienen poco impacto en las prácticas docentes y no pueden aplicar sus competencias y habilidades al mejoramiento de la enseñanza. Su función está algunas veces restringida *de facto* al control de los abusos al interior de la profesión.

**Instrumentos.** La evidencia con respecto a la práctica de los docentes se recopila mediante la observación de una sesión de clases, seguida de una entrevista con el maestro. Las críticas a este enfoque incluyen: (i) el hecho de que una observación de aula aislada podría no ser suficiente como para obtener una visión equitativa y rigurosa de las habilidades y conocimientos del maestro, y (ii) en la entrevista, los maestros se focalizan en reaccionar a las críticas de los inspectores en lugar de analizar sus necesidades de mejoramiento en particular. Todo el procedimiento no parece ofrecer mucho espacio a la autoevaluación y a la reflexión por parte de los maestros con respecto a su propia práctica y desempeño.

**Criterios.** Se observan y califican los aspectos tanto ‘pedagógicos’ como ‘administrativos’, pero sin referencia a un marco que defina lo que es una ‘buena’ enseñanza. Las inquietudes son numerosas. La naturaleza de las diferentes habilidades ‘pedagógicas’ evaluadas como también su ponderación en la apreciación general del maestro siguen estando en gran medida a discreción de cada inspector. Esto refuerza las evaluaciones subjetivas y los resultados impredecibles y aleatorios a expensas de la equidad y rigurosidad del proceso. Los maestros informan que no saben cómo y basado en qué criterios están siendo evaluados. Los criterios más objetivos y bien comprendidos utilizados para evaluar a los maestros son aquellos ‘administrativos’, tales como la puntualidad y la asistencia. En consecuencia, la calificación obtenida por un maestro suele estar principalmente determinada por su calificación en cuanto a certificación (es decir, el resultado del examen de ingreso).

**Consecuencias.** Las consecuencias de la evaluación docente en la carrera de un maestro son limitadas, excepto en los casos de graves faltas de conducta. Los sueldos de los docentes están determinados por una escala salarial única en la cual la progresión depende de los años de servicio junto con las calificaciones iniciales y el examen de ingreso. El compromiso con el trabajo es rara vez reconocido y valorado al igual que el mérito, el desempeño destacado o las iniciativas que buscan mejorar el aprendizaje de los estudiantes. Además, no existe ningún vínculo con las activida-

des de desarrollo profesional y estas últimas son muy limitadas y están desconectadas de las deficiencias identificadas en los docentes. El proceso de evaluación no ofrece oportunidades de auto-reflexión con respecto a las prácticas docentes ni de aprendizaje mutuo entre pares y comprenden poca asesoría y *coaching*.

Para mayor información:

- *Haut Conseil de l'évaluation de l'école* (2003): [http://cisad.adc.education.fr/hcee/documents/rapport\\_annuel\\_2003.pdf](http://cisad.adc.education.fr/hcee/documents/rapport_annuel_2003.pdf)
- Informes de las inspecciones generales: <http://lesrapports.ladocumentationfrancaise.fr/BRP/054004446/0000.pdf>
- Ministerio de Educación: <http://www.education.gouv.fr/cid263/l-evaluation-des-personnels.html>

Para otros sistemas de evaluación del desempeño docente:

- La Provincia de Alberta en Canadá: <http://www.education.alberta.ca/departement/policy/k12manual/section2/teacher.aspx>
- El Estado de Iowa en Estados Unidos: <http://www.iowa.gov/educate/content/view/1450/1617>

## REFERENCIAS

- Aaronson, D.; Barrow, L. y Sander, W. (2007) “Teachers and Student Achievement in the Chicago Public High Schools”, *Journal of Labor Economics*, Vol. 25, No. 1, pp. 95-135.
- American Federation of Teachers (2001) “Beginning Teacher Induction: The Essential Bridge”, *Educational Issues Policy Brief No. 13*, AFT, 2001.
- American Federation of Teachers y National Education Association (2008) *A Guide to Understanding National Board Certification*, AFT and NEA, Washington, DC.
- Anderson, L. y Pellicer, L. (2001) *Teacher Peer Assistance and Review*, Corwin Press.
- Avalos, B. y Assael, J. (2006) “Moving from resistance to agreement: The case of the Chilean teacher performance evaluation”, *International Journal of Educational Research*, Vol. 45, No. 4-5, pp. 254-266.
- Beck, R.; Livne, N. y Bear, S. (2005) “Teachers’ self-assessment of the effects of formative and summative electronic portfolios on professional development”, *European Journal of Teacher Education*, Vol. 28, No. 3, pp. 221-244.
- Bolino, M. y Turnley, W. (2003) “Counter normative impression management, likeability, and performance ratings: the use of intimidation in an organizational setting”, *Journal of Organizational Behavior*, Vol. 24, No. 2, pp. 237-250.
- Bond, L.; Smith, T.; Baker, W. y Hattie, J. (2000) “The Certification System of the National Board for Professional Teaching Standards: A Construct and Consequential Validity Study”, NBPTS, 2000.
- Borman, G. y Kimball, S. (2005) “Teacher Quality and Educational Equality: Do Teachers with Higher Standards-Based Evaluation Ratings Close Student Achievement Gaps?”, *The Elementary School Journal*, Vol. 106, No. 1, pp. 3-20.
- Braun, H. (2005) “Using Student Progress to Evaluate Teachers: A Primer on Value-Added Models”, *Educational Testing Service (ETS)*, 2005.
- Casson, H. Jr (2007) “Reducing Teacher Moral Hazard in the U.S. Elementary and Secondary Educational System through Merit-pay: An Application of the Principal - Agency Theory”, *Forum for Social Economics*, Vol. 36, No. 2, pp. 87-95.
- Campbell, D.; Melenzyer, B.; Nettles, D. y Wyman, R. (2000) *Portfolio and performance assessment in teacher education*, Needham Heights, MA, Allyn & Bacon.
- Cavalluzzo, L. (2004) “Is National Board Certification An Effective Signal of Teacher Quality?”, *The CNA Corporation*, Alexandria, Virginia, 2004.
- Center for Assessment and Evaluation of Student Learning (2004) “Using Student Tests to Measure Teacher Quality”, *CAESL Assessment Brief No. 9*.

- Center for Teaching Quality (2008) "Measuring What Matters: The Effects of National Board Certification on Advancing 21<sup>st</sup> Century Teaching and Learning", CTQ, 2008.
- Cohen, C. y King Rice, J. (2005) "National Board Certification as Professional Development: Design and Cost", NBPTS, 2005.
- Corcoran, T. (2007) "Teaching Matters: How State and Local Policymakers Can Improve the Quality of Teachers and Teaching", Consortium for Policy Research in Education (CPRE) Policy Briefs RB-48.
- Danielson, C. (2001) "New Trends in Teacher Evaluation", *Educational Leadership*, Vol. 58, No. 5, pp. 12-15.
- Danielson, C. (1996, 2007) *Enhancing Professional Practice: A Framework for Teaching*, 1<sup>st</sup> and 2<sup>nd</sup> editions, Association for Supervision and Curriculum Development (ASCD), Alexandria, Virginia.
- Danielson, C. y McGreal, T. (2000) *Teacher Evaluation to Enhance Professional Practice*, Association for Supervision and Curriculum Development (ASCD), Alexandria, Virginia.
- Darling, L. (2001) "Portfolio as practice: the narratives of emerging teachers", *Teaching and Teacher Education*, Vol. 17, pp. 107-121.
- Darling-Hammond, L.; Pecheone, R. y Stansbury, K. (2004) "Beginning Teacher Quality: What Matters for Student Learning?", Research proposal from Stanford University to the Carnegie Corporation of New York, available at [www.pacttpa.org/\\_files/Publications\\_and\\_Presentations/Carnegie\\_grant\\_proposal.doc](http://www.pacttpa.org/_files/Publications_and_Presentations/Carnegie_grant_proposal.doc).
- Day, C. y Gu, Q. (2007) "Variations in the conditions for teachers' professional learning and development: sustaining commitment and effectiveness over a career", *Oxford Review of Education*, Vol. 33, No. 4, pp. 423-443.
- Department of Education, Science and Training (2007) "Performance-based rewards for teachers", DEST Research Papers, 2007.
- Elmore, R. (2000) *Building a New Structure for School Leadership*, Albert Shanker Institute, Winter 2000.
- Figlio, D. y Kenny, L. (2007), "Individual teacher incentives and student performance", *Journal of Public Economics*, Vol. 91, No. 5-6, pp. 901-914.
- Freund, M.; Kane Russell, V. y Kavulic, C. (2005) "A Study of the Role of Mentoring in Achieving Certification by the National Board for Professional Teaching Standards", NBPTS, 2005.
- Goe, L. (2007) "The Link Between Teacher Quality and Student Outcomes: A Research Synthesis", National Comprehensive Center for Teacher Quality, 2007.
- Goldhaber, D. y Anthony, E. (2007) "Can Teacher Quality Be Effectively Assessed? National Board Certification As a Signal of Effective Teaching", *The Review of Economics and Statistics*, Vol. 89, No. 1, pp. 134-150.
- Greenfield, W. (1995) "Toward a Theory of School Administration: the Centrality of Leadership", *Educational Administration Quarterly*, Vol. 31, No. 1, pp. 61-85.
- Halverson, R.; Kelley, C. y Kimball, S. (2004) "Implementing Teacher Evaluation Systems: How Principals Make Sense of Complex Artifacts to Shape Local Instructional Practice" in *Educational Administration, Policy and Reform: Research and Measurement Research and Theory in Educational Administration*, Vol. 3, W.K. Hoy and C.G. Miskel (Eds.) Greenwich, CT: Information Age Press.
- Hanushek, E. (2004) "Does School Accountability Lead to Improved Student Performance?", NBER Working Papers n° 10591.
- Hanushek, E. (1992) "The Trade-Off between Child Quantity and Quality", *Journal of Political Economy*, Vol. 100, No. 1, pp. 84-117.
- Hanushek, E. (1986) "The Economics of Schooling: Production and Efficiency in Public Schools", *Journal of Economic Literature*, Vol. 24, No. 3, pp. 1141-1177.
- Hanushek, E.; Kain, J.; O'Brien, D. y Rivkin, S. (2005) "The Market for Teacher Quality", NBER Working Papers n° 11154.
- Harris, D. y Sass, T. (2007) "The Effects of NBPTS-Certified Teachers on Student Achievement", Center for Analysis of Longitudinal Data in Education Research (CALDER), Working Paper n° 4.
- Haut Conseil de l'évaluation de l'école (2003) *Rapport annuel*, HCéé, 2003.
- Heneman, H. y Milanowski, A. (2003) "Continuing Assessment of Teacher Reactions to a Standards-Based Teacher Evaluation System", *Journal of Personnel evaluation in Education*, Vol. 17, No. 2, pp. 173-195.
- Heneman, H.; Milanowski, A. y Kimball, S. (2007) "Teacher Performance Pay: Synthesis of Plans, Research, and Guidelines for Practice", Consortium for Policy Research in Education (CPRE) Policy Briefs RB-46.
- Heneman, H.; Milanowski, A.; Kimball, S. y Odden, A. (2006) "Standards-Based Teacher Evaluation as a Foundation for Knowledge- and Skill-Based Pay", Consortium for Policy Research in Education (CPRE) Policy Briefs RB-45.
- Hess, F. y West, M. (2006) "A Better Bargain: Overhauling Teacher Collective Bargaining for the 21<sup>st</sup> Century", Cambridge, MA: Program on Education Policy and Governance, Harvard University.
- Holland, P. (2005) "The Case for Expanding Standards for Teacher Evaluation to Include an Instructional Supervision Perspective", *Journal of Personnel Evaluation in Education*, Vol. 18, No. 1, pp. 67-77.
- Ingvarson, L.; Kleinhenz, E. y Wilkinson, J. (2007) *Research on Performance Pay for Teachers*, Australian Council for Educational Research (ACER), 2007.
- Interstate New Teacher Assessment and Support Consortium (1992) "Model Standards for Beginning Teacher Licensing, Assessment and Development: A Resource for State Dialogue", INTASC, Council of Chief State School Officers (CCSSO), 1992.
- Jacob, B. (2004) "Accountability, Incentives and Behavior: The Impact of High-Stakes Testing in the Chicago Public Schools", *Journal of Public Economics*, Vol. 89, No. 5-6, pp. 761-796.
- Jacob, B. y Lefgren, L. (2008) "Can Principals Identify Effective Teachers? Evidence on Subjective Performance Evaluation in Education", *Journal of Labor Economics*, Vol. 26, No. 1, pp. 101-136.
- Jacob, B. y Lefgren, L. (2005b) "What Do Parents Value in Education: an Empirical Investigation of Parents' Revealed Preferences for Teachers", NBER Working Paper n° 11494.
- Jacob, B. y Lefgren, L. (2005a) "Principals as Agents: Subjective Performance Measurement in Education", NBER Working Papers n° 11463.
- Jacob, B. y Levitt, S. (2003) "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating", *The Quarterly Journal of Economics*, Vol. 118, No. 3, pp. 843-878.
- Jacobs, C.; Martin, S. y Otieno, T. (2008) "A Science Lesson Plan Analysis Instrument for Formative and Summative Program Evaluation of a Teacher Education Program", *Science Teacher Education* (Articles online in advance of print).
- Jun, M.-K.; Anthony, R.; Achrazoglu, J. y Coghill-Behrends, W. (2007) "Using ePortfolio for the Assessment and Professional Development of Newly Hired Teachers", *TechTrends*, Vol. 51, No. 4, pp. 45-50.
- Kane, T. y Staiger, D. (2002) "Volatility in School Test Scores: Implications for Test-Based Accountability Systems", *Brooking Papers on Education Policy*, Washington, DC.
- Kennedy, M. (2005) *Inside Teaching*, Harvard University Press, London, England, 2005.
- Kimball, S. (2002) "Analysis of Feedback, Enabling Conditions and Fairness Perceptions of Teachers in Three School Districts with New Standards-Based Evaluation Systems", *Journal of Personnel Evaluation in Education*, Vol. 16, No. 4, pp. 241-268.
- Kimball, S.; Milanowski, T. y McKinney, S. (2007) "Implementation of Standards-Based Principal Evaluation in One School District: First Year Results From Randomized Trial", Paper presented at the annual conference of the American Educational Research Association, available at <http://cpre.wceruw.org/publications/KimballMilanowskiMcKinney.pdf>.
- Klecker, B. (2000) "Content validity of preservice teacher portfolios in a standards-based program", *Journal of Instructional Psychology*, Vol. 27, No. 1, pp. 35-38.
- Kleinhenz, E. y Ingvarson, L. (2004) "Teacher Evaluation Uncoupled: A Discussion of Teacher Evaluation Policies and Practices in Australian States and Their Relation to Quality Teaching and Learning", *Research Papers in Education*, Vol. 19, No. 1, pp. 31-49.
- Leigh, A. (2007) "Estimating Teacher Effectiveness From Two-Year Changes in Students' Test Scores", *Research School of Social Sciences*, Australian National University, available at <http://econrssh.anu.edu.au/~aleigh/pdf/TQPanel.pdf>.

- Levin, J. (2003) "Relational incentive contracts", *American Economic Review*, Vol. 93, No. 3, pp. 835-57.
- Lustick, D. y Sykes, G. (2006) "National Board Certification as Professional Development: What are Teachers Learning?", *Education Policy Analysis Archives*, Vol. 14, No. 5.
- MacLeod, B. (2003) "Optimal contracting with subjective evaluation", *American Economic Review* 93, No. 1, pp. 216-240.
- Mansvelde-Longayroux, D.; Beijaard, D. y Verloop, N. (2007) "The portfolio as a tool for stimulating reflection by student teachers", *Teaching and Teacher Education*, Vol. 23, No. 1, pp. 47-62.
- Margo, J.; Benton, M.; Withers, K. y Sodha, S. (2008) *Those Who Can?*, Institute for Public Policy Research (IPPR) Publications, 2008.
- Marshall, K. (2005) "It's Time to Rethink Teacher Supervision and Evaluation", *Phi Delta Kappan*, Vol. 86, No. 10, pp. 727-735.
- McColskey, W. y Stronge, J. (2005) "A Comparison of National Board Certified Teachers and non-National Board Certified Teachers: Is there a difference in teacher effectiveness and student achievement", NBPTS, 2005.
- Milanowski, A. (2007) "Performance Pay System Preferences of Students Preparing to Be Teachers", *American Education Finance Association*, 2007.
- Milanowski, A. (2004) "The Relationship Between Teacher Performance Evaluation Scores and Student Achievement: Evidence From Cincinnati", *Peabody Journal of Education*, Vol. 79, No. 4, pp. 33-53.
- Milanowski, A. y Heneman, H. (2001) "Assessment of Teacher Reactions to a Standards-Based Teacher Evaluation System: A Pilot Study", *Journal of Personnel Evaluation in Education*, Vol. 15, No. 3, pp. 193-212.
- Milanowski, A. y Kimball, S. (2003) "The Framework-Based Teacher Performance Assessment Systems in Cincinnati and Washoe", CPRE Working Paper Series TC-03-07.
- Ministerial Council on Education, Employment Training and Youth Affairs (2003) "A National Framework for Professional Standards for Teaching", MCEETYA, Carlton South, Australia, 2003.
- Mizala, A. y Romaguera, P. (2004) "School and teacher performance incentives: The Latin American experience", *International Journal of Educational Development*, Vol. 24, No. 6, pp. 739-754.
- Muñoz, M. y Chang, F. (2007) "The Elusive Relationship Between Teacher Characteristics and Student Achievement Growth: A Longitudinal Multilevel Model for Change", *Journal of Personnel Evaluation in Education*, Vol. 20, No. 3-4, pp. 147-164.
- Nabors Oláh, L.; Lawrence, N. y Riggan, M. (2008) "Learning to learn from benchmark assessment data: How teachers analyze results", Paper presented at the Annual Meeting of the American Educational Research Association, New York, 2008, available at [http://www.cpre.org/images/stories/cpre\\_pdfs/aera2008\\_olah\\_lawrence\\_riggan.pdf](http://www.cpre.org/images/stories/cpre_pdfs/aera2008_olah_lawrence_riggan.pdf).
- National Board for Professional Teaching Standards (2007) "A Research Guide on National Board Certification of Teachers", NBPTS, Arlington, VA, 2007.
- Odden, A. y Kelley, C. (2002) *Paying Teachers for What They Know and Do: New and Smarter Compensation Strategies to Improve Schools*, Corwin Press, Thousand Oaks, California, 2002.
- Office for Standards in Education (2006) "The logical chain: continuing professional development in effective schools", OFSTED Publications n° 2639, United Kingdom, 2006.
- Organisation for Economic Co-Operation and Development (2008) *Improving School Leadership*, OECD, Paris, 2008.
- Organisation for Economic Co-Operation and Development (2005) *Teachers Matter: Attracting, Developing and Retaining Effective Teachers*, OECD, Paris, 2005.
- Ovando, M. y Ramirez, A Jr (2007) "Principals' Instructional Leadership Within a Teacher Performance Appraisal System: Enhancing Students' Academic Success", *Journal of Personnel Evaluation in Education*, Vol. 20, No. 1-2, pp. 85-110.
- Pecheone, R. y Chung, R. (2006) "Evidence in Teacher Education: The Performance Assessment for California Teachers (PACT)", *Journal of Teacher Education*, Vol. 57, No. 1, pp. 22-36.
- Peterson, K. (2000) *Teacher Evaluation: A Comprehensive Guide to New Directions and Practices*, 2<sup>nd</sup> edition, Thousand Oaks, CA: Corwin Press.
- Peterson, K.; Wahlquist, C. y Bone, K. (2000) "Student Surveys for Teacher Evaluation", *Journal of Personnel Evaluation in Education*, Vol. 14, No. 2, pp. 135-153.
- Peterson, K.; Wahlquist, C.; Esparza Brown, J. y Mukhopadhyay, S. (2003) "Parents Surveys for Teacher Evaluation", *Journal of Personnel Evaluation in Education*, Vol. 17, No. 4, pp. 317-330.
- Peterson, K. (2006) "Teacher Pay Reform Challenges States", *Stateline.org: where policy and politics news click*, available at <http://www.stateline.org/live/ViewPage.action?siteNodeId=136&languageId=1&contentId=93346>.
- Petty, T. (2002) "Identifying the Wants and Needs of North Carolina High School Mathematics Teachers for Job Success and Satisfaction", NBPTS, 2002.
- Ping Yan Chow, A.; King Por Wong, E.; Seeshing Yeung, A. y Wan Mo, K (2002) "Teachers' Perceptions of Appraiser-Appraisee Relationships", *Journal of Personnel Evaluation in Education*, Vol. 16, No. 2, pp. 85-101.
- Pochard, M. (2008) *Libre vert sur l'évolution du métier d'enseignant, Rapport au ministre de l'Éducation nationale, La Documentation française, Collection des rapports officiels*, 2008.
- Popham, J. (1997) "Consequential validity: Right Concern - Wrong Concept", *Educational Measurement: Issues and Practice*, Vol. 16, No. 2, pp. 9-13.
- Robinson, V. (2007) "School Leadership and Student Outcomes: Identifying What works and Why", Australian Council for Educational Leaders, ACEL Monograph Series No. 41.
- Sanders, W.; Ashton, J. y Wright, P. (2005) "Comparison of the Effects of NBPTS Certified Teachers with Other Teachers on the Rate of Student Academic Progress", NBPTS, 2005.
- Smith, T.; Gordon, B.; Colby, S. y Wang, J. (2005) "An Examination of the Relationship Between Depth of Student Learning and National Board Certification Status", Office for Research on Teaching, Appalachian State University.
- Stronge, J. y Tucker, P. (2003) *Handbook on Teacher Evaluation: Assessing and Improving Performance*, Eye On Education Publications, 2003.
- Stronge, J.; Ward, T.; Tucker, P. y Hindman, J. (2007) "What is the Relationship Between Teacher Quality and Student Achievement? An Exploratory Study", *Journal of Personnel Evaluation in Education*, Vol. 20, No. 3-4, pp. 165-184.
- Strudler, N. y Wertzell, K. (2008) "Costs and Benefits of Electronic Portfolios in Teacher Education: Faculty Perspectives", *Journal of Computing in Teacher Education*, Vol. 24, No. 4, pp. 135-142.
- Training and Development Agency for Schools (2007a) "Models Performance Management Policy for Schools", TDA, United Kingdom, 2007.
- Training and Development Agency for Schools (2007b) "Professional Standards for Teachers: Why Sit Still in Your Career?", TDA, United Kingdom, 2007.
- Tucker, P.; Stronge, J. y Gareis, C. (2002) *Handbook on teacher portfolios for evaluation and professional development*, Larchmont, NY, Eye on Education.
- UNESCO (2007) *Evaluación del Desempeño y Carrera Profesional Docente: Una panorámica de América y Europa*, Oficina Regional de Educación para América Latina y el Caribe, UNESCO Santiago, 2007.
- Vandervoort, L.; Amrein-Beardsley, A. y Berliner, D. (2004) "National Board Certified Teachers and Their Students' Achievement", *Education Policy Analysis Archives*, Vol. 12, No. 46.
- Weingarten, R. (2007) "Using Student Test Scores to Evaluate Teachers: Common Sense or Nonsense?", United Federation of Teachers, available at [http://www.uft.org/news/randi/ny\\_times/UFT\\_WMM\\_Mar07\\_v32.pdf](http://www.uft.org/news/randi/ny_times/UFT_WMM_Mar07_v32.pdf).
- Wertzell, K. y Strudler, N. (2006) "Costs and Benefits of Electronic Portfolios in Teacher Education: Student Voices", *Journal of Computing in Teacher Education*, Vol. 22, No. 3, pp. 69-78.
- Xin, T.; Xu, Z. y Tatsuoka, K. (2004) "Linkage Between Teacher Quality, Student Achievement, and Cognitive Skills: A Rule-Space Model", *Studies in Educational Evaluation*, Vol. 30, pp. 205-223.



Programa de Promoción de la Reforma Educativa en América Latina y el Caribe  
Partnership for Educational Revitalization in the Americas

El Programa de Promoción de la Reforma Educativa en América Latina y el Caribe es un proyecto conjunto del Diálogo Interamericano, con sede en Washington, y la Corporación de Investigaciones para el Desarrollo (CINDE), con sede en Santiago de Chile.

Desde su creación en 1995, el PREAL ha tenido como objetivo central contribuir a mejorar la calidad y equidad de la educación en la región mediante la promoción de debates informados sobre temas de política educacional y reforma educativa, la identificación y difusión de buenas prácticas y la evaluación y monitoreo del progreso educativo.

La ejecución de las actividades se realiza a través de Centros Asociados de Investigación y Políticas Públicas en diversos países de la región y comprenden la realización de estudios, la organización de debates y la promoción de diálogos públicos sobre temas de política educacional y reforma educativa.

Las actividades del PREAL son posibles gracias al apoyo de la United States Agency for International Development (USAID), el Banco Interamericano de Desarrollo (BID), el Banco Mundial, la International Association for the Evaluation of Educational Achievement (IEA), The Tinker Foundation, GE Foundation, entre otras.



Inter-American Dialogue • 1211 Connecticut Ave. N.W. Suite 510  
Washington, D.C. 20036 U.S.A. • Tel: (202) 822-9002  
Fax: (202)822-9553 • E-mail: [iad@thedialogue.org](mailto:iad@thedialogue.org)  
Internet: [www.thedialogue.org](http://www.thedialogue.org) & [www.preal.org](http://www.preal.org)

CINDE • Santa Magdalena 75, Piso 10 • Oficina 1002 • Providencia  
Santiago, Chile • Tel: (56-2) 334-4302  
Fax: (56-2) 334-4303 • E-mail: [infopreal@preal.org](mailto:infopreal@preal.org)  
Internet: [www.preal.org](http://www.preal.org)

