

PROFESORES A PRUEBA

Claves para una evaluación docente exitosa



Yyannú Cruz-Aguayo | Diana Hincapié | Catherine Rodríguez

PROFESORES A PRUEBA

PROFESORES A PRUEBA

Claves para una evaluación docente exitosa

Yyannú Cruz-Aguayo | Diana Hincapié | Catherine Rodríguez

**Catalogación en la fuente proporcionada por la
Biblioteca Felipe Herrera del
Banco Interamericano de Desarrollo**

Cruz-Aguayo, Yyannú.

Profesores a prueba: claves para una evaluación docente exitosa / Yyannú Cruz-Aguayo,
Diana Hincapié, Catherine Rodríguez.

p. cm. — (Monografía del BID ; 749)

Incluye referencias bibliográficas.

978-1-59782-368-5 (Rústica)

978-1-59782-369-2 (Digital)

1. Teachers-Rating of-Latin America. 2. Teacher effectiveness-Latin America-Evaluation.
3. Teaching-Latin America-Evaluation. I. Hincapié, Diana. II. Rodríguez, Catherine.
III. Banco Interamericano de Desarrollo. División de Educación. IV. Título. V. Serie.
IDB-MG-749

Palabras clave: Calidad educativa, Calidad docente, Evaluación docente

Clasificación JEL: I21, I28, J24, O57

Copyright © [2020] Banco Interamericano de Desarrollo.

Esta obra se encuentra sujeta a una licencia Creative Commons IGO 3.0 Reconocimiento-
NoComercial-SinObrasDerivadas (CC-IGO 3.0 BY-NC-ND) (<http://creativecommons.org/licenses/by-nc-nd/3.0/igo/legalcode>) y puede ser reproducida para cualquier uso no-comercial otorgando el reconocimiento respectivo al BID. No se permiten obras derivadas.

Cualquier disputa relacionada con el uso de las obras del BID que no pueda resolverse amistosamente se someterá a arbitraje de conformidad con las reglas de la CNUDMI (UNCITRAL). El uso del nombre del BID para cualquier fin distinto al reconocimiento respectivo y el uso del logotipo del BID, no están autorizados por esta licencia CC-IGO y requieren de un acuerdo de licencia adicional.

Note que el enlace URL incluye términos y condiciones adicionales de esta licencia.

Las opiniones expresadas en esta publicación son de los autores y no necesariamente reflejan el punto de vista del Banco Interamericano de Desarrollo, de su Directorio Ejecutivo ni de los países que representa.



CONTENIDO

Lista de cuadros	vii
Lista de gráficos	ix
Agradecimientos	xi
Sobre las autoras	xiii
1 La evaluación docente: identificando la efectividad y apoyando el desarrollo de los maestros	1
2 Por qué es importante evaluar a los docentes y cómo hacerlo bien	9
Instrumentos disponibles para la evaluación docente	13
Pruebas estandarizadas de aprendizajes a los estudiantes	13
Medidas de objetivos de aprendizaje	15
Observaciones estandarizadas en clase	16
Portafolio docente	18
Encuestas a estudiantes y a otros actores	18
Pruebas estandarizadas a los docentes	19
Estudios que analizan diferentes instrumentos	19
Evaluaciones de impacto de sistemas de evaluación docente	23
Impactos en la composición del cuerpo docente, en la retención de los docentes y en las habilidades pedagógicas de los profesores	23
Impactos en los resultados académicos de los estudiantes	26
3 Cómo se evalúa a los docentes en los sistemas educativos exitosos	31
Instrumentos utilizados en las evaluaciones docentes	35
Pesos asignados dentro de la evaluación docente	40
Consecuencias ligadas a los resultados dentro de la evaluación	42

4 Situación de los sistemas de evaluación docente en América Latina y el Caribe	47
Los estándares de desempeño docente en la región.	53
La reglamentación de la profesión en la región	56
Evaluaciones para el ingreso a la carrera docente	60
Programas de inducción para docentes principiantes y evaluaciones para superar el período de prueba	66
Evaluaciones obligatorias periódicas.	69
Evaluaciones para ascensos horizontales.	86
5 Cinco claves para una evaluación docente exitosa	95
Referencias bibliográficas	103

Apéndice en línea disponible en la página de publicaciones del BID

LISTA DE CUADROS

Cuadro 3.1	Instrumentos utilizados en las evaluaciones docentes en algunos sistemas educativos exitosos seleccionados	37
Cuadro 3.2	Pesos asignados a cada instrumento utilizado en las evaluaciones docentes de algunos estados y distritos de Estados Unidos	41
Cuadro 3.3	Consecuencias ligadas a los resultados de la evaluación docente en los sistemas educativos exitosos seleccionados	43
Cuadro 4.1	Evaluaciones mencionadas de manera explícita en los estatutos de cada sistema educativo	59
Cuadro 4.2	Número y tipo de instrumentos utilizados en los concursos docentes en la región	65
Cuadro 4.3	Implementación efectiva y multiplicidad de instrumentos de la evaluación obligatoria de acuerdo con la periodicidad de su aplicación	72
Cuadro 4.4	Instrumentos utilizados en países con evaluaciones docentes obligatorias con múltiples instrumentos	74
Cuadro 4.5	Actores que participan en la evaluación obligatoria periódica en países con múltiples instrumentos	75
Cuadro 4.6	Peso promedio asignado a cada instrumento en los países que utilizan instrumentos múltiples en las evaluaciones obligatorias periódicas de docentes	78
Cuadro 4.7	Instrumentos definidos para las evaluaciones de ascenso horizontales en los países donde esta posibilidad existe y la legislación incluye aspectos meritocráticos	82

LISTA DE GRÁFICOS

Gráfico 2.1	Teoría de cambio de las evaluaciones docentes	11
Gráfico 4.1	Tipos de evaluaciones docentes analizadas en el estudio	48
Gráfico 4.2	Avance de los estándares de desempeño docente en América Latina y el Caribe	54
Gráfico 4.3	Reglamentación de la carrera docente en la región a lo largo del tiempo	57
Gráfico 4.4	Obligatoriedad, centralización y niveles educativos requeridos en los concursos docentes de la región	62
Gráfico 4.5	Tipos de evaluaciones del período de prueba a las cuales están sujetos los docentes principiantes	68
Gráfico 4.6	Evaluaciones obligatorias periódicas en la región	71
Gráfico 4.7	Resultados de la evaluación docente periódica obligatoria en países seleccionados	81
Gráfico 4.8	Consecuencias de los resultados obtenidos en las evaluaciones obligatorias periódicas de acuerdo con la periodicidad de su implementación	84
Gráfico 4.9	Posibilidades de ascensos horizontales en las carreras docentes de la región y sus requisitos, según los estatutos docentes	88
Gráfico 4.10	Frecuencia en el uso de instrumentos en evaluaciones de ascensos horizontales meritocráticos en las carreras docentes de la región	92

AGRADECIMIENTOS

Profesores a prueba: claves para una evaluación docente exitosa es una publicación de la División de Educación del Sector Social del Banco Interamericano de Desarrollo (BID).

Agradecemos especialmente a Norbert Schady, asesor económico del Sector Social, y a Gregory Elacqua, economista principal de la División de Educación, quienes proporcionaron orientación y retroalimentación a lo largo de este proyecto. Agradecemos también el apoyo de Emiliana Vegas, exjefa de la División de Educación, Sabine Rieble-Auborg, jefa interina de la División de Educación; Marcelo Cabrol, gerente del Sector Social; Ana María Ibáñez, asesora económica principal de la Vicepresidencia de Sectores y Conocimiento, y Ana María Rodríguez, vicepresidenta de Sectores y Conocimiento.

Nuestros colegas de la División de Educación nos facilitaron los contactos con las autoridades educativas de los países y brindaron retroalimentación útil en diversas etapas de la producción de este estudio.

Varios investigadores participaron de la elaboración de los documentos de apoyo que se utilizaron para este trabajo: María Camila Rivera apoyó la elaboración de un estudio sobre sistemas de evaluación docente en América Latina y el Caribe; James Wyckoff y Verónica Katz prepararon un estudio sobre sistemas de evaluación docente en Estados Unidos, y Phillippa Codringley y Paul Crisp, del Centre for the Use of Research and Evidence in Education (CUREE) produjeron un trabajo sobre los sistemas de evaluación docente en países de alto desempeño educativo. Diana Paredes brindó asistencia en varias etapas del proceso, particularmente en la edición y diagramación del texto.

Agradecemos a los expertos en Educación y a las autoridades educativas de los sistemas educativos de América Latina y el Caribe, que conversaron con nosotras, resolvieron nuestras dudas, y validaron gran parte de la información recolectada.

El libro recibió comentarios de revisores externos anónimos. Agradecemos también las observaciones de los comentaristas y participantes en el Diálogo Regional de Políticas de la Red de Educación del BID, especialmente de Santiago Cueto y Ximena Dueñas.

La edición estuvo a cargo de Claudia M. Pasquetti y de Alejandra Abrodos.

Los comentarios y opiniones que se expresan en esta publicación pertenecen a las autoras, y no reflejan en forma alguna la visión del Banco Interamericano de Desarrollo ni de su Directorio Ejecutivo.

SOBRE LAS AUTORAS

Yyannú Cruz-Aguayo se desempeña como economista senior en la División de Mercados Laborales del BID. Es especialista en evaluaciones de impacto de programas sociales y recientemente co-lideró la agenda de investigación sobre el impacto de la calidad docente en los resultados del aprendizaje en la escolaridad temprana. Ha trabajado en el Banco Mundial, en el Banco Nacional de México y en la Secretaría de Hacienda de dicho país. Es autora y co-autora de artículos académicos (en revistas como *Quarterly Journal of Economics*) y de capítulos de libros sobre la efectividad de los programas sociales y el desarrollo infantil temprano, entre otros temas. Actualmente colabora en el desarrollo de instrumentos de préstamo para países miembros del Banco. Tiene una maestría y un doctorado de la Universidad de Maryland, College Park. Su agenda de investigación en la División de Mercados Laborales se centra en la efectividad de intervenciones laborales para poblaciones desfavorecidas.

Diana Hincapié es economista de la División de Educación del BID. Dirige proyectos de investigación sobre la calidad de la educación, la política docente, el desarrollo de habilidades a lo largo de la vida, el desarrollo infantil temprano y la jornada escolar extendida. Trabaja en el diseño y la implementación de evaluaciones de impacto de programas y políticas educativas, y de préstamos y operaciones de asistencia técnica para los países de América Latina y el Caribe. Es co-autora de la publicación insignia del BID de 2017 (DIA) *Aprender mejor: políticas públicas para el desarrollo de habilidades* y del libro *Profesión: profesor en América Latina. ¿Por qué se perdió el prestigio docente y cómo recuperarlo?*, así como autora de artículos académicos. Ha trabajado en el Departamento de Investigación del BID, la Unidad de Pobreza y Género del Banco Mundial, y el Centro de Estudios sobre Desarrollo Económico (CEDE) de la Universidad de Los Andes en Colombia. Es economista de la Universidad de Los Andes (Colombia) y magíster en Economía de la misma universidad. Asimismo, tiene un doctorado en Políticas Públicas y Administración Pública de la Universidad George Washington.

Catherine Rodríguez es economista experta en evaluaciones de impacto de programas educativos, en el diseño de política docente y en temas asociados con las consecuencias del conflicto armado en la población civil. Es autora de libros, capítulos de libros y

artículos académicos en revistas indexadas nacionales e internacionales. Actualmente se desempeña como investigadora de la Facultad de Economía y el Centro de Estudios sobre Desarrollo Económico (CEDE) de la Universidad de los Andes. Ha sido profesora asociada de esa misma facultad y consultora del BID, del Banco de Desarrollo de América Latina (CAF) y del Banco Mundial, entre otras instituciones. Es economista de la Universidad del Rosario (Colombia), tiene una maestría en Política Económica y un doctorado en Economía de Boston University.



CAPÍTULO 1

LA EVALUACIÓN DOCENTE: IDENTIFICANDO LA EFECTIVIDAD Y APOYANDO EL DESARROLLO DE LOS MAESTROS

Una educación de calidad es indispensable para promover el bienestar de las personas, el crecimiento económico y la reducción de las desigualdades sociales de los países. Por lo tanto, no es de extrañar que los gobiernos de América Latina y el Caribe hayan realizado esfuerzos importantes en las últimas décadas para aumentar la cobertura educativa. La mayoría de los países de América Latina ha alcanzado una cobertura universal en la educación primaria, mientras que el 77% de los niños de 4 a 5 años y el 80% de los jóvenes de 15 a 17 años asisten a la escuela (BID-CIMA, 2019). Con el aumento en cobertura, los años promedio de educación se incrementaron de prácticamente cero a comienzos del siglo XX a nueve en la actualidad (Busso, Cristiá, Hincapié, Messina y Ripani, 2017).

A pesar de estos innegables avances, las habilidades de los estudiantes muestran que aún hay grandes rezagos en la calidad educativa de los sistemas de la región. Por ejemplo, solo aproximadamente un tercio de los alumnos de cuarto grado que participaron en las pruebas del Tercer Estudio Regional Comparativo y Explicativo (TERCE) y un tercio de los alumnos de 15 años que participaron en las pruebas del Programa para la Evaluación Internacional de Alumnos (PISA, por sus siglas en inglés) lograron alcanzar un nivel mínimo de competencias en matemáticas y lectura (Busso et al., 2017). Por ello, es necesario seguir avanzando en la implementación de políticas que busquen mejorar la calidad educativa en nuestros países.

Entre los esfuerzos para lograrlo, se deben incluir políticas que abarquen los diferentes aspectos que determinan dicha calidad. Por el lado de la oferta educativa, sin duda, una

de las variables fundamentales es la calidad de los docentes. Las investigaciones en el área educativa han demostrado que los docentes son el insumo escolar más importante para promover el desarrollo de los aprendizajes. Algunos análisis de la literatura han concluido que las intervenciones que involucran a los maestros son particularmente exitosas en mejorar la calidad educativa (véanse, por ejemplo: Ganimian y Murnane, 2016; Kremer, Brannen y Glennerster, 2013; Evans y Popova, 2016; McEwan, 2015). Más aún, existe evidencia de que el impacto de un buen docente va más allá de los aprendizajes en la escuela, llegando a influir en otros ámbitos, como, por ejemplo, en las tasas de graduación, el aumento de la probabilidad de acceder y culminar estudios superiores, en los ingresos laborales e incluso en la reducción del embarazo adolescente (Jackson, 2018; Chetty, Friedman y Rockoff, 2014b).

Por este motivo, no es sorprendente que parte de los factores que se encuentran detrás del éxito de los sistemas educativos de algunos países como Finlandia, Japón, Dinamarca, Suiza y República de Corea sea tener una política coherente de selección, capacitación y desarrollo de la fuerza docente que ha llevado a que la profesión sea altamente respetada y valorada. Por el contrario, como lo evidenciaron Elacqua, Hincapié, Vegas y Alfonso (2018), la valoración y el prestigio de la profesión docente en América Latina cayó en las últimas décadas del siglo XX. Además, las habilidades pedagógicas y los conocimientos disciplinares de los docentes en la región están por debajo de lo que sugieren los estándares internacionales. Por ejemplo, los docentes de la región dedican 20% menos del tiempo efectivo en clase de lo que estos recomiendan. Es decir, en América Latina, las diferencias en tiempos de enseñanza efectiva de los docentes implican que los estudiantes de la región reciben en promedio un día menos de clase a la semana (Bruns y Luque, 2015). De manera similar, en lo que respecta a los temas disciplinares, pruebas en Perú, Chile y México y estudios internacionales aplicados a los propios docentes indican que sus conocimientos de matemáticas son insatisfactorios (Elacqua et al., 2018).

Para revertir esta situación y lograr una verdadera transformación de la calidad educativa en la región, es indispensable que nuestros gobiernos asuman compromisos serios para desarrollar e implementar una política sistémica que permita mejorar el prestigio de la profesión y la calidad de los docentes. Esta política debe incluir aspectos a lo largo de la carrera docente, como los siguientes: i) la atracción de mejores candidatos a carreras de Educación y Pedagogía; ii) la mejora de la calidad y pertinencia de la formación inicial; iii) la selección de candidatos a la profesión docente que esté basada en el mérito; iv) un marco que regule la profesión de manera clara, transparente y meritocrática; v) una política de remuneración competitiva que genere incentivos a la excelencia; y vi) una evaluación de los docentes que asegure el mejoramiento continuo (Elacqua et al., 2018; Bruns y Luque, 2015; García et al., 2014).

Todas estas políticas deben tener como objetivo desarrollar y potenciar los conocimientos disciplinares y las habilidades pedagógicas de los docentes. Sin embargo, esta no es tarea sencilla. Tanto para países desarrollados como en desarrollo, la evidencia sugiere que incluso en un mismo país (con un mismo marco regulatorio y administrativo) hay variaciones significativas en la calidad de los maestros. Más aún, estas diferencias están

presentes dentro de una misma escuela, en donde las prácticas pedagógicas y las habilidades para promover el aprendizaje de los estudiantes son bastante diferentes entre profesores. Esto lo han demostrado estudios recientes para Estados Unidos (Hanushek y Rivkin, 2010; Chetty, Friedman y Rockoff, 2014a); y para varios países de América Latina (Araujo et al., 2016; Bruns y Luque, 2015).

Las evaluaciones docentes pueden ayudar a identificar las diferencias de desempeño entre los profesores. Además, el uso adecuado de sus resultados puede otorgar la información necesaria para aprovechar al máximo sus fortalezas, buscar superar las falencias, y potenciar la excelencia en la profesión. **No obstante, para que las evaluaciones docentes cumplan estos propósitos y sean realmente aceptadas, implementadas y utilizadas en cualquier sistema educativo se requiere que sean válidas y confiables. Esto significa que la información que surja de ellas debe ser objetiva, estar realmente asociada a su labor diaria y reflejar aspectos fundamentales dentro de los procesos de aprendizaje de los estudiantes, de modo que permita que tanto los propios docentes como las autoridades educativas entiendan y reconozcan cuáles son las fortalezas y debilidades en la labor del docente.** Más aún, solo si los resultados de las evaluaciones son utilizados adecuadamente en programas de desarrollo profesional y/o en incentivos en la carrera —que incluyan bonificaciones, ascensos o remoción de las aulas dependiendo del caso— se puede mejorar la calidad de la enseñanza y el aprendizaje de los estudiantes.

Distintos sistemas educativos en el mundo están llevando a cabo grandes esfuerzos en este aspecto y transformando los sistemas de evaluación docente. Diversos factores pueden explicar la evolución que este tipo de evaluaciones han tenido en los últimos años. El bajo impacto de los indicadores tradicionalmente usados para medir el impacto de los docentes sobre los aprendizajes de los estudiantes —como el nivel de educación o experiencia del docente— ha incentivado la búsqueda de medidas más directamente relacionadas con su productividad y efectividad. La introducción de políticas laborales que buscan incentivar un mayor esfuerzo y asignar una mayor responsabilidad del desempeño de los alumnos a los propios docentes también puede explicar este cambio. Finalmente, la búsqueda de la excelencia académica en los sistemas educativos y la evidencia de la importancia de la labor docente en ella es otro factor que se encuentra detrás del énfasis reciente en torno a los sistemas de evaluación docente.

Los países de América Latina no son ajenos a esta transformación y en los últimos años varios de ellos han puesto en marcha iniciativas para mejorar los procesos de evaluación docente en el sistema educativo público. Incluso hay algunos modelos de evaluación docente en la región que pueden considerarse de “segunda generación” y que se encuentran en distintos niveles de desarrollo e implementación.¹ Aprovechando esta coyuntura y

¹ Como se explicará en detalle más adelante, este tipo de evaluaciones, consistentes con las múltiples labores que llevan a cabo los docentes, se caracterizan por contar con una multiplicidad de instrumentos y evaluadores con el objetivo de entender en detalle los distintos aspectos de los procesos de enseñanza, y buscando asegurar las características de validez y confiabilidad antes mencionadas.

en un contexto en donde es indispensable hacer un salto en la efectividad de los docentes para mejorar la calidad de la educación en la región, **este estudio presenta algunas claves para lograr el diseño de una evaluación docente exitosa.**

El estudio se divide en cinco capítulos, incluida esta introducción. El segundo capítulo explica por qué es importante evaluar a los docentes, describe diferentes instrumentos que pueden ser incluidos dentro de los sistemas de evaluación y los aspectos más relevantes a tener en cuenta sobre su diseño y resultados. El tercer capítulo presenta un análisis de los procesos de evaluación docente existentes en 12 sistemas que, de acuerdo con pruebas estandarizadas internacionales, ofrecen a sus estudiantes una elevada calidad educativa. El cuarto capítulo detalla el estado actual de los sistemas de evaluación docente en 19 sistemas educativos de América Latina y el Caribe. Finalmente, el quinto capítulo expone algunas recomendaciones de política para el diseño de un sistema de evaluación exitoso.

Asegurar que los sistemas de evaluación docente sean confiables y válidos no es trivial. Al momento de su diseño e implementación, las autoridades educativas se enfrentan a una multiplicidad de posibilidades, que incluyen la selección de instrumentos, la decisión de cómo deben o pueden implementarse y el peso que se le asignará a cada uno dentro del puntaje final, entre otros elementos. Igualmente, una vez que los procesos finalizan, debe decidirse cómo entregar la información recolectada a los docentes y, lo que es más importante aún, cómo utilizarla para asegurar la mejora de la labor pedagógica.

Vale la pena aclarar que un paso previo al diseño del sistema de evaluación y a la elección de los instrumentos a ser utilizados tiene que ver con la definición de lo que debe saber y debe saber hacer un buen docente en cada país. Estas características, que en general se encuentran definidas en los estándares de desempeño docente, constituyen una herramienta esencial para guiar y fundamentar todos los ejes dentro de la profesión docente, incluido el de la evaluación. Por supuesto, no es posible definir bajo un marco común lo que se espera de un docente de excelencia y que este pueda ser aplicado de forma específica en todos los países de la región. Incluso, dentro de cada país, lo que se espera o requiere de los docentes puede variar según el contexto en el que esté trabajando, por ejemplo: si es una escuela rural, una escuela con población indígena o una escuela multigrado.

A pesar de que los estándares de desempeño constituyen un aspecto fundamental dentro de un sistema de evaluación docente, no son tratados en profundidad en este estudio. Sin embargo, es claro que se espera que los docentes cuenten con conocimientos disciplinares y pedagógicos adecuados a su materia y grado de enseñanza, sobre sus estudiantes y su proceso de aprendizaje, sobre su contexto y sobre el uso en el aula de materiales pedagógicos y recursos tecnológicos, de tal forma que aseguren una enseñanza efectiva.² Esto es

² La transformación tecnológica está modificando las prácticas de aprendizaje y enseñanza en el mundo y es un aspecto que debe tenerse en cuenta en la región. La tecnología puede complementar el trabajo de los maestros para fortalecer el aprendizaje de diferentes maneras (Muralidharan, Singh y Ganimian, 2019; Lai et al., 2012; Banerjee et al., 2007). Por ejemplo, un *software* de aprendizaje dinámico puede

precisamente lo que el uso adecuado de la información que surge de sistemas de evaluación válidos y confiables, basados en estándares claros, podría ayudar a asegurar y lo que se analiza en detalle en el libro.

El segundo capítulo resume información relevante para los formuladores de políticas sobre algunos de los instrumentos más comunes para evaluar a los docentes y describe los aspectos más destacados que se conocen con respecto a su diseño y a sus resultados.

La elección de qué instrumentos incluir en una evaluación docente no es trivial. Dentro de la información plasmada en el capítulo resaltan dos programas de investigación extensos y recientes en el tema. El primero, el proyecto de Medidas de Enseñanza Efectiva (MET, por sus siglas en inglés, Measures of Effective Teaching) de 2013, permitió concluir que, a través de una combinación de instrumentos, que incluyen la observación en clase, medidas de valor agregado y encuestas estudiantiles, es posible identificar a los docentes más efectivos y que logran tener mayor impacto en el aprendizaje de sus estudiantes. Sin embargo, las conclusiones obtenidas en el segundo estudio (Stecher et al., 2018) dejan en evidencia que llevar a la práctica una evaluación efectiva es difícil y costoso. Solo es posible obtener resultados positivos cuando la implementación de múltiples instrumentos de evaluación es cuidadosa y tiene en cuenta cada uno de los eslabones previamente mencionados.

El tercer capítulo presenta un análisis de los procesos de evaluación docente en 12 sistemas educativos que se caracterizan por ofrecer a sus estudiantes una alta calidad educativa. Aunque no es conveniente ni deseable implementar o copiar sistemas de evaluación que funcionan en contextos distintos, la revisión de experiencias internacionales puede ofrecer ideas y aprendizajes valiosos para entender cómo estos países están sorteando las difíciles decisiones en torno a los procesos de evaluación docente. Los sistemas incluidos en la revisión son Estonia, Ontario (Canadá), Nueva Zelanda, Reino Unido, República de Corea y Singapur; y en Estados Unidos se analizan los sistemas de Denver Public Schools (DPS), District of Columbia Public Schools (DCPS), New Haven Public Schools (NHPS), New York City Public Schools (NYCPS) y los estados de Nuevo México y Tennessee. Aunque los aprendizajes que se pueden obtener del estudio de estos casos son múltiples, la revisión se enfocó en tres aspectos particulares que, además de ser fundamentales dentro de los procesos de evaluación, permiten una comparación puntual a lo largo de estos 12 sistemas: i) el tipo de instrumentos de evaluación escogidos dentro de cada sistema; ii) el peso que se le asigna a cada uno dentro del puntaje total de la evaluación docente y iii) el uso que cada sistema educativo le da a los resultados de la evaluación.

De este análisis de los sistemas internacionales, surgen tres conclusiones clave. Primero, las evaluaciones docentes implementadas en estos sistemas educativos —excepto en tres casos particulares— abarcan una multiplicidad de instrumentos que buscan reflejar

hacer que los alumnos aprendan a su propio ritmo y puede adaptarse dinámicamente a su conocimiento. Este tipo de *software* puede permitir que los maestros conozcan las fortalezas y debilidades de sus estudiantes y puedan brindar un servicio más personalizado.

la complejidad de la labor docente. En todos los casos se incluye la observación de clase, la diferencia crucial entre ellos se encuentra en los detalles y la rigurosidad con respecto a cómo esta se implementa. Para los estados y distritos analizados dentro de Estados Unidos, es común también la inclusión de medidas de rendimiento estudiantil tanto objetivas —basadas en medidas de valor agregado o percentiles de progreso académico del estudiante— como subjetivas, basadas en objetivos de aprendizaje. Algunos sistemas han agregado también encuestas a estudiantes o padres de familia, y en una menor proporción se observa la utilización de instrumentos como la autoevaluación o el portafolio de la labor docente. Segundo, sobre la base del análisis de los casos de Estados Unidos, el instrumento que recibe el mayor peso dentro de la evaluación docente es el de la observación en clase, seguido por instrumentos que evalúan el desempeño estudiantil. Tercero, la gran mayoría de los sistemas vincula los resultados de las evaluaciones a bonificaciones o a promociones laborales dentro de las carreras docentes. Llama la atención que son pocos los sistemas analizados que tienen explícitamente definidos aspectos de formación en servicio y posibilidades de permanencia. Dos excepciones claras son las de Singapur y República de Corea, donde el tema de la formación en servicio constituye una parte fundamental dentro del proceso de evaluación docente implementado.

El cuarto capítulo detalla el estado actual de los sistemas de evaluación docente en 19 sistemas educativos de la región: provincia de Buenos Aires (Argentina), Belice, Bolivia, Santa Catarina (Brasil), Chile, Colombia, Costa Rica, Ecuador, El Salvador, Guatemala, Honduras, Jamaica, México, Panamá, Paraguay, Perú, República Dominicana, Trinidad y Tobago, y Uruguay.

La revisión comienza con el diagnóstico de dos temas íntimamente ligados a la evaluación: la existencia de estándares de desempeño docente y el nivel de meritocracia definido por los estatutos docentes en cada sistema. El primero de estos temas, como ya se recalcó, es un eslabón crucial para poder diseñar e implementar cualquier política en torno a los docentes, particularmente en los ejes de formación y evaluación. El segundo aspecto permite entender cuáles son los objetivos de política explícitos en cada caso analizado, centrando la atención en el tema de la evaluación y en las posibilidades de ascensos dentro de las carreras. Después de entender estos dos aspectos, el capítulo sistematiza las distintas evaluaciones que hoy en día se implementan en la región en tres momentos de la carrera docente pública o estatal: i) evaluaciones para el ingreso; ii) evaluaciones en el período de prueba y iii) evaluaciones para docentes en ejercicio. Dentro de esta última etapa se analizan dos tipos de evaluaciones específicas: aquellas obligatorias que se llevan a cabo con una periodicidad definida y con posibilidad de desvinculación en caso de un bajo desempeño reiterado, y aquellas que se realizan para acceder a ascensos dentro de la carrera. Para cada una de estas evaluaciones, el capítulo resume las principales características en torno a su existencia, el tipo de instrumentos utilizados, el peso que se le asigna a cada uno de ellos, los encargados de su implementación, y el uso de resultados por parte de las autoridades educativas y los propios maestros.

Aunque son muchos los avances y también las diferencias que se observan en los sistemas de evaluación de la región, el análisis del capítulo 4 permite destacar tres conclusiones fundamentales en torno a las evaluaciones docentes en América Latina y el Caribe. **Primero**, existen discrepancias notables en cuanto a los niveles de meritocracia definidos dentro de los estatutos que rigen la profesión docente en la región, que incluyen aspectos relacionados con los procesos para el ingreso, los ascensos y la salida de la carrera, entre otros. Al enfocar el análisis en la mención o no de la evaluación docente en el estatuto, se encontró que existen países en donde ni siquiera se hace referencia a ella, mientras que hay otros países cuyos estatutos tienen de manera explícita y detallada los cuatro tipos de evaluaciones consideradas en el estudio. **Segundo**, la situación *de facto* no necesariamente refleja la situación *de jure* en lo que se refiere a la implementación y al uso de resultados de la evaluación docente. Así, aunque la legislación defina la obligatoriedad y las consecuencias en términos laborales de las evaluaciones docentes en las distintas etapas de la profesión de manera detallada, estas evaluaciones no siempre se aplican, o si se aplican, no se implementan de la forma en que están en la legislación o no todas las consecuencias que de ellas se deben derivar efectivamente ocurren. **Tercero**, a pesar de estas debilidades, es claro que un gran número de países de la región está haciendo esfuerzos valiosos que buscan mejorar la atracción, retención y profesionalización de la fuerza docente a través de procesos de evaluación. De hecho, en un número no despreciable de países, se aplican o se están comenzando a aplicar evaluaciones que se pueden catalogar como de *segunda generación*. Además, en algunos países la evaluación ha venido acompañada de investigaciones acerca de los procesos e impactos de la implementación de estos sistemas de evaluación, lo cual permite optimizar los instrumentos y protocolos utilizados en ellos. Es precisamente este tipo de información la que permitirá a los directivos y autoridades del sector tomar decisiones informadas que se encaminen hacia la mejora de la calidad de la educación brindada.

Finalmente, **el quinto capítulo concluye el estudio con algunas recomendaciones de política y las claves para tener una evaluación docente exitosa**. Se espera que estas sugerencias, junto con los demás análisis presentados, resulten útiles para guiar y potenciar los esfuerzos llevados a cabo en la región en el tema de evaluación docente, y para alcanzar una mejora continua de la calidad de la enseñanza para todos los estudiantes de nuestra región.



CAPÍTULO 2

POR QUÉ ES IMPORTANTE EVALUAR A LOS DOCENTES Y CÓMO HACERLO BIEN

La investigación en el área de educación ha aumentado y se ha enriquecido de manera considerable en las últimas décadas. Las herramientas y las técnicas estadísticas para el análisis de los datos han mejorado. Asimismo, la cantidad y la calidad de los datos y la información disponible también han avanzado de forma notable. En la actualidad no solo se cuenta con mejores medidas del nivel de aprendizaje de los estudiantes, sino que, además, los investigadores en el área de educación están recolectando información mucho más precisa acerca de las habilidades de los docentes en su conocimiento disciplinar y en el pedagógico.

Estos avances han permitido demostrar que los docentes son el factor escolar más importante para los aprendizajes de los estudiantes en la escuela. También resulta claro que existen diferencias significativas en la calidad docente, y que incluso estas diferencias muchas veces son mayores al interior de una escuela que entre distintas escuelas. Por ejemplo, al analizar el uso del tiempo en clase por parte de los docentes en América Latina, Bruns y Luque (2015) hallaron que mientras un docente en una escuela de México dedicaba el 80% del tiempo a instrucción efectiva otro docente de la misma escuela no destinaba ni un solo minuto de la clase a ello. De hecho, de acuerdo con los autores, estas diferencias son tan comunes que la dispersión en el uso del tiempo entre el “mejor” y el “peor” profesor de la misma escuela explica en promedio dos tercios de la variación total de los aprendizajes entre los estudiantes de esa escuela.

En ese contexto, no sorprenden los resultados de otros estudios según los cuales alumnos de una misma escuela asignados a docentes con distintos niveles de efectividad alcanzan niveles de aprendizaje muy diferentes. El análisis de Hanushek y Rivkin (2010) para Estados Unidos y más recientemente lo demostrado por Araujo et al. (2016) para Ecuador,

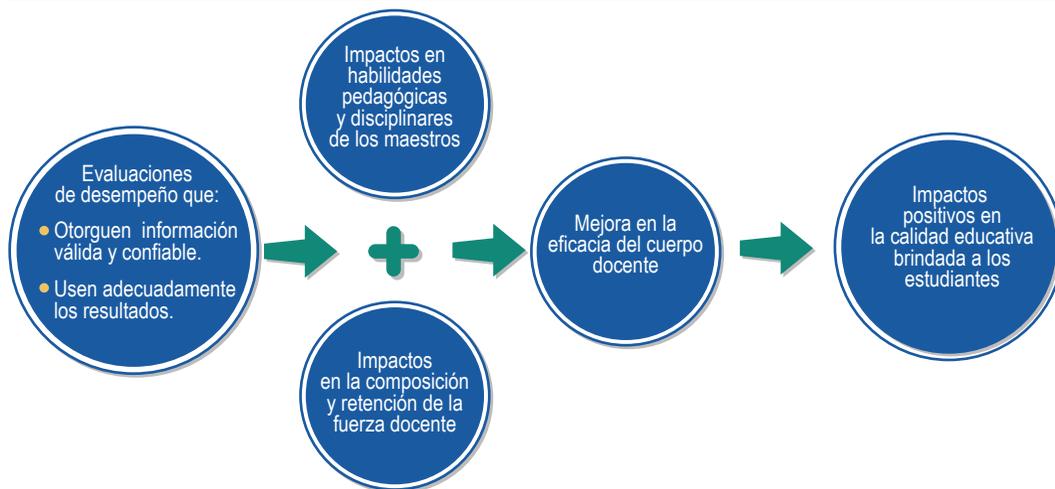
sugieren que alumnos asignados a profesores de alto desempeño dominan el contenido enseñado cerca de un 50% más que uno asignado a un docente de bajo desempeño, lo que equivale a poco más de un año escolar. Estudios recientes demuestran que los docentes impactan también en las habilidades socioemocionales de sus estudiantes medidas a través de *proxies* observables en el comportamiento de los jóvenes, como ausencias, suspensiones y graduación de la secundaria (Jackson, 2018), y también en mediciones directas de estas habilidades, incluida la determinación, la forma de pensar y el esfuerzo (Kraft, 2019), aunque la dificultad para medir estas habilidades ha influido en el hecho de que algunos estudios no encuentren impactos significativos. Incluso se sabe con certeza que los impactos de un buen docente no solo abarcan los aprendizajes de sus estudiantes en la escuela, sino que repercuten también en resultados de más largo plazo como los relacionados con las decisiones de educación superior, con aspectos laborales y hasta con temas como el embarazo adolescente (Chetty, Friedman y Rockoff, 2014b). Es importante tener en cuenta que, si bien idealmente desearíamos tener docentes que impactaran positivamente en todas estas habilidades y variables, la literatura sugiere que los que tienen impactos en aprendizajes no necesariamente van a tenerlo en las habilidades socioemocionales y viceversa. Este es un aspecto importante a tener en cuenta a la hora de medir y evaluar la efectividad de los docentes.

Todo esto ha provocado que hoy las políticas que buscan atraer y formar docentes efectivos, así como retenerlos en las aulas, estén en el centro del debate educativo. Una de estas políticas es la evaluación docente. Como quedará evidenciado en los capítulos siguientes, son varios los gobiernos de distintos países del mundo —incluso de América Latina— que han llevado a cabo esfuerzos significativos en tiempo y en dinero a fin de mejorar este aspecto específico de la carrera docente. Estos esfuerzos, aunque han sido inmensos, son relativamente recientes, por lo cual la investigación sobre este tema es escasa, incipiente y, además, está concentrada en casos particulares. Esto no quiere decir, sin embargo, que no existan enseñanzas importantes de las cuales los formuladores de políticas, los académicos y, en general, aquellos interesados en el tema puedan aprender. Este capítulo busca resumir las principales conclusiones que se pueden extraer de estudios causales que analizan aspectos relacionados con el diseño, la implementación y los impactos que puede generar la aplicación de la evaluación docente en los sistemas educativos.

La primera pregunta que surge al analizar el tema es cómo podría la evaluación docente atraer, formar y retener en las aulas a los mejores docentes, y a través de esto mejorar el aprendizaje de los estudiantes. El gráfico 2.1 presenta la teoría de cambio, resumiendo los mecanismos a través de los cuales esto puede llegar a suceder.¹ Todo parte del diseño y la

¹ Para esta teoría de cambio se asume que los docentes con mayores habilidades pedagógicas y disciplinares efectivamente las despliegan en las aulas y de esta forma pueden mejorar el aprendizaje de sus estudiantes. Asimismo, se asume que los docentes cuentan con los recursos pedagógicos y el material de apoyo necesarios para poder ser efectivos en el aula.

GRÁFICO 2.1 | Teoría de cambio de las evaluaciones docentes



Fuente: Elaboración propia.

implementación de evaluaciones de desempeño docente basadas en instrumentos que otorguen información válida y confiable. Un instrumento (o un conjunto de instrumentos) de evaluación docente otorga información **válida** si los resultados que surgen de su implementación permiten identificar, mediante investigaciones rigurosas, docentes altamente efectivos o que tienen un impacto en el aprendizaje de sus estudiantes. Un instrumento (o un conjunto de instrumentos) otorga información **confiable** cuando los resultados que se obtienen de cada docente reflejan su desempeño típico en clase y no dependen del día en particular en que la información fue recolectada o de la persona que estuvo a cargo de esa recolección. Estas dos cualidades son fundamentales para que la autoridad educativa pueda entender cuáles son las fortalezas y las debilidades de sus docentes y pueda llevar adelante las acciones necesarias para potenciarlas o superarlas, respectivamente. Además, estas cualidades también son esenciales para que el cuerpo docente acepte la evaluación, confíe en ella y actúe de acuerdo con sus resultados. Por tal motivo, al realizar las evaluaciones es clave producir información y análisis que sustenten la confiabilidad y la validez de estas, según lo recomiendan los estándares internacionales sobre evaluaciones educativas (AERA, APA y NCME, 2014).

La segunda característica necesaria dentro de la teoría de cambio (reflejada en el segundo punto del primer círculo de la cadena) es el uso que se les da a los resultados de la evaluación. Aunque algunas de las investigaciones que se detallan más adelante demuestran que la sola participación en el proceso de evaluación puede traer impactos positivos en las prácticas docentes (Taylor y Tyler, 2012), los expertos argumentan que el mecanismo a través del cual evaluaciones cuidadosamente diseñadas e implementadas pueden mejorar la calidad docente depende del uso que se les dé a los resultados. De acuerdo con

la Organización para la Cooperación y el Desarrollo Económicos (OCDE) este uso puede tener dos objetivos: i) mejorar las prácticas y las habilidades pedagógicas y/o disciplinares a partir del diagnóstico y la vinculación a programas de desarrollo profesional diseñados para superar los resultados; ii) mejorar la composición y la motivación de la fuerza docente por medio del otorgamiento de bonificaciones, reconocimientos especiales o ascensos para aquellos docentes con resultados excelentes o excluir al docente del sistema —o al menos retirarlo de las aulas— en los casos en que muestre de manera consistente que no cumple con las condiciones requeridas por la profesión (OCDE, 2009 y 2013).

El problema que surge en relación con estos dos objetivos (plasmados en el segundo eslabón del gráfico 2.1) es que son difíciles de lograr con una única herramienta de evaluación. Para poder detectar los aspectos por mejorar de las prácticas y el conocimiento pedagógico y disciplinar, y reconocer la excelencia docente, es necesario que los docentes estén completamente abiertos a revelar sus prácticas y logros y dispuestos a compartirlos con las autoridades. Esto permitiría entender las necesidades de formación de la fuerza docente y buscar superarlas. Sin embargo, si los resultados de la evaluación también están atados a aspectos de la carrera como son los salarios, los ascensos e incluso la estabilidad laboral, puede que tanto la aceptación de la puesta en marcha de los procesos de evaluación como la actitud positiva y abierta de los docentes y evaluadores hacia ella se vean comprometidas.

La solución a este problema no es sencilla. De hecho, como lo demuestran análisis previos (OCDE, 2013; Steinberg y Donaldson, 2016; Wyckoff y Katz, 2018; CUREE, 2018) y, como quedará claro en la revisión de las experiencias internacionales y en las de América Latina y el Caribe reseñadas en los capítulos siguientes, los países han optado por distintas soluciones. Algunos sistemas educativos enfocan el uso de los resultados en alcanzar solo uno de los dos objetivos; otros implementan múltiples tipos de evaluaciones a fin de lograr un objetivo distinto con cada evaluación, y otros, en cambio, evaden esta complejidad y buscan conseguir ambos objetivos por medio de una sola evaluación. Por supuesto, no existe una respuesta única y correcta a esta disyuntiva y la solución adecuada para cada país dependerá del contexto del sistema educativo y la fuerza docente.

Los últimos dos eslabones del gráfico 2.1 cierran la teoría de cambio detrás de las evaluaciones docentes. Como es posible observar, se asume que el uso adecuado de los resultados obtenidos en las evaluaciones docentes impacta en los niveles de eficiencia del cuerpo docente tanto en términos pedagógicos como disciplinares (dependiendo por supuesto del tipo de instrumentos utilizados). Esto, finalmente, terminará impactando en el conocimiento que adquieren los alumnos en las aulas, mejorando la calidad educativa.

Independientemente de las decisiones respecto del uso de los resultados, es crucial que el sistema asegure que la información brindada sea válida y confiable. En este aspecto también existe una multiplicidad de decisiones que deben tomar los formuladores de políticas, ya que hay una gama amplia de instrumentos disponibles para evaluar a los docentes. La siguiente sección resume algunos de los instrumentos más comunes y lo que dicen sobre su efectividad las investigaciones al respecto. Por su parte, la última sección de este capítulo

resume lo que investigaciones rigurosas han encontrado en cuanto a los impactos del diseño y de la puesta en marcha de evaluaciones docentes en algunos sistemas educativos.

INSTRUMENTOS DISPONIBLES PARA LA EVALUACIÓN DOCENTE

Al ser la docencia una actividad compleja, una evaluación holística de la labor docente requiere una combinación de distintos instrumentos. Esto implica que al momento de diseñar los sistemas de evaluación docente las autoridades educativas deben escoger los instrumentos que serán utilizados, pero también deben definir los pesos que se le asignarán a cada uno de ellos. Estas decisiones tendrán repercusiones en los costos del sistema, el nivel de aceptación por parte de los diferentes actores y el éxito de su implementación. Por ejemplo, asignar mayor peso a un determinado indicador cuantitativo puede generar el aumento en la presión alrededor de los resultados obtenidos de este y puede hacer más probable la distorsión de procesos que, se supone, debería monitorear (Campbell, 1979).

Aunque la revisión detallada de los instrumentos disponibles, sus características y los requisitos para su correcta aplicación está fuera del alcance de este trabajo, es valioso mencionar los instrumentos utilizados más comúnmente y lo que las principales investigaciones han evidenciado acerca de ellos.

Pruebas estandarizadas de aprendizajes a los estudiantes

El objetivo final del diseño correcto, la implementación y el uso de resultados de evaluaciones de desempeño docente es mejorar los aprendizajes de los estudiantes en el sistema. No sorprende, por lo tanto, que diversos sistemas educativos utilicen instrumentos basados en exámenes estandarizados a los estudiantes dentro de la evaluación docente. De hecho, el 97% de los docentes incluidos en la muestra del Estudio Internacional sobre Enseñanza y Aprendizaje (TALIS, por sus siglas en inglés) de 2013 reporta que alguna medida de desempeño de sus estudiantes es utilizada como parte de las evaluaciones docentes que se llevan a cabo en su escuela (Smith y Kubacka, 2017). De forma consistente con esta evidencia, la revisión de Steinberg y Donaldson (2016) para Estados Unidos demuestra que esta información también es ampliamente utilizada por las autoridades educativas de ese país en sus sistemas de evaluación.

Entre los instrumentos que utilizan este tipo de información los dos más comunes son las medidas de eficiencia docente de valor agregado y los percentiles de progreso académico del estudiante. En el primer caso se usan los resultados anteriores de exámenes de conocimiento realizados a los alumnos para estimar, controlando por un conjunto de características personales y de la escuela, el resultado esperado para cada uno de ellos. Las diferencias entre el valor esperado en la prueba de cada estudiante y el que efectivamente obtuvo el alumno se promedian para todos los estudiantes de cada docente, lo cual da como resultado el valor agregado del docente. En el segundo caso, las medidas

de crecimiento en percentil comparan el crecimiento en el logro de un estudiante con el crecimiento de otros estudiantes que en años anteriores tuvieron medidas de desempeño similares. Luego se mide la efectividad del docente estimando el crecimiento en el percentil promedio de sus alumnos.

Los defensores de la inclusión de las pruebas estandarizadas de conocimiento a los estudiantes como un instrumento dentro de las evaluaciones docentes les atribuyen tres aspectos positivos:

- i. Aunque son conscientes de que los exámenes de conocimiento no capturan todos los objetivos que se busca alcanzar en las escuelas y que en consecuencia deberían ser parte de la labor docente, destacan que estas pruebas sí miden el aprendizaje de sus estudiantes, que es uno de los objetivos más importantes.
- ii. La literatura de impacto ha demostrado que medidas de valor agregado presentes de los docentes son buenos predictores de medidas de valor agregado en el futuro (Kane et al., 2013; Araujo et al., 2016). Por ejemplo, para el caso de Ecuador, Araujo et al. (2016) encuentran que, además de la experiencia y calidad de las interacciones de los docentes con sus alumnos dentro del aula de clase, el valor agregado también es una medida que predice los aprendizajes estudiantiles.
- iii. Las medidas de valor agregado docente están correlacionadas con aspectos de desarrollo individual de sus estudiantes en el largo plazo como la asistencia a la educación superior, los salarios que devengan cuando tienen 28 años y el embarazo adolescente (Chetty et al., 2014a y 2014b).

Por otro lado, existen argumentos *en contra* del uso de pruebas estandarizadas de conocimiento a los estudiantes como parte de la evaluación docente, que se detallan a continuación.

1. El primero de ellos señala que las pruebas estandarizadas no pueden identificar todo el conjunto de resultados que se busca alcanzar en el proceso de enseñanza-aprendizaje.
2. El segundo estipula que, en parte, el desempeño de los estudiantes está fuera del alcance absoluto del docente y depende de los propios estudiantes y de su motivación e interés al presentar las pruebas. Dado que los docentes no pueden efectivamente impactar sobre todos los resultados en las pruebas a estudiantes, no es conveniente asignar un peso importante a los resultados en dichas pruebas.
3. El tercer argumento indica que, incluso si se acepta el aprendizaje disciplinar como una medida fundamental a alcanzar dentro del sistema educativo, es preciso considerar que las medidas de valor agregado no suelen estar disponibles para todos los docentes dentro de los sistemas educativos sino únicamente para aquellos cuyos alumnos presentan los exámenes estandarizados en los cuales se deben basar. Este no es un punto menor ya que, por ejemplo, en países desarrollados la información solo está

disponible para alrededor del 20% de los docentes y en América Latina el porcentaje es aún menor (Wyckoff y Katz, 2018).²

4. El cuarto argumento apunta que, como toda medida, su confiabilidad depende de la validez de los exámenes de conocimiento que responden los estudiantes, factor que no siempre está asegurado (Lockwood et al., 2007; Papay, 2011).
5. El quinto establece que cuando hay incentivos vinculados con los resultados de la evaluación de aprendizajes de los estudiantes el diseño debe ser cuidadoso para asegurar que no se conviertan en estrategias de enseñanza solo para los exámenes (“teaching to the test”) o que incluso puedan estimular la copia o la trampa³ para aumentar los puntajes y obtener así los incentivos (Koretz, 2015; Stecher, 2002).
6. El sexto argumento indica que el cálculo de las medidas de valor agregado es complejo, incluso metodológicamente, y requiere de decisiones que pueden alterar los resultados, como por ejemplo la elección del tipo de controles de características socioeconómicas de los estudiantes y de las escuelas que se incluyen en sus estimaciones (Donaldson y Papay, 2015). Más aún, hay dudas respecto del grado de confiabilidad del uso de estas medidas, ya que su correlación anual es relativamente baja y solo con promedios a lo largo de cinco años aumenta a niveles aceptables (Kane, Kerr y Pianta, 2014).
7. Por último, la séptima crítica está vinculada con el momento en el que medidas de efectividad docente basadas en estos resultados se pueden estimar y el uso que de ellas puede surgir. En general, estas medidas se calculan y se difunden al finalizar el año académico, o incluso después. Más importante aún, otorgan poca información que les permita a los docentes y a la autoridad educativa entender en qué aspecto puede y/o debe mejorar cada docente. De acuerdo con Smith y Kubacka (2017), los docentes consideran que sistemas de evaluación que se basan de manera significativa en este instrumento tienen un impacto limitado en sus estrategias de enseñanza y lo ven como un proceso meramente administrativo y poco formativo.

Si bien muchas de estas críticas pueden ser válidas en ciertos contextos, la clave para una evaluación docente exitosa está en encontrar la forma de aprovechar las ventajas que estas pueden ofrecer, e incorporar en el diseño mismo de la evaluación formas de mitigar los elementos negativos que usualmente se les atribuyen.

Medidas de objetivos de aprendizaje

El instrumento basado en medidas de objetivos de aprendizaje está compuesto por los objetivos que los docentes establecen al comienzo del año escolar respecto de los logros

² Para superar, en parte, la crítica respecto de que estas medidas están disponibles solo para un porcentaje pequeño de docentes algunos sistemas educativos han optado por utilizar medidas de valor agregado promedio a nivel de escuela. Sin embargo, en estos casos la distancia entre la efectividad del docente y medidas de resultado agregadas es aún mayor, lo cual le resta conveniencia al uso de estrategias como esta.

³ Por ejemplo, los docentes podrían excluir a algunos estudiantes de menor rendimiento académico de las pruebas para intentar obtener un puntaje mejor.

académicos que esperan que sus alumnos alcancen. El superior jerárquico del docente revisa y aprueba estos objetivos y al finalizar el año académico evalúa, junto con el docente, el progreso efectivamente alcanzado por los alumnos. En función de dicho progreso, el docente recibe una calificación de su desempeño. La ventaja más evidente de este tipo de instrumento en comparación con aquellos basados en pruebas estandarizadas a los estudiantes es que puede establecerse para todos los docentes del sistema educativo, al margen de la materia o el grado en que los docentes dictan sus clases. Sin embargo, su nivel de estandarización y objetividad es bajo. Aunque, en general, están sujetos a objetivos o competencias definidos previamente por cada sistema, para cada área del conocimiento y grado educativo, en última instancia dependen de la autonomía que cada docente tiene en su definición y en la rigurosidad de los superiores al evaluar el nivel en que estos objetivos han sido cumplidos.⁴ De hecho, Lachlan-Haché et al. (2015), en uno de los estudios más completos sobre el tema, mencionan que la literatura especializada permite concluir que la implementación de este instrumento no solo es subjetiva sino que demanda mucho tiempo por parte de los docentes y superiores y está muy débilmente correlacionada con los resultados de los estudiantes en pruebas estandarizadas. A pesar de esto, como quedará claro en las revisiones de experiencias internacionales, es uno de los instrumentos más utilizados en gran parte de los sistemas educativos de Estados Unidos y el peso que se le asigna dentro de la calificación final que se les da a los docentes es significativo.⁵

Observaciones estandarizadas en clase

Las observaciones en clase de la labor del docente son el instrumento de evaluación docente más antiguo de todos. Incluso antes de la implementación de sistemas de evaluación con múltiples instrumentos y con distintas consecuencias en todo el mundo, los docentes siempre han sido observados de una forma u otra por superiores jerárquicos. Lo que ha cambiado en los últimos años es la estructura en la que estas observaciones se llevan a cabo y las consecuencias que puede tener esta acción. En algunos sistemas educativos se basan en rúbricas de observación de clase diseñadas específicamente para cada sistema. De manera ideal estas rúbricas deberían estar basadas en estándares de desempeño docente que a su vez definan lo que se espera que un buen maestro deba saber y saber hacer en cada país. Aunque hay rúbricas ya predefinidas y ampliamente utilizadas en investigaciones educativas, como Framework for Teaching (FFT) (Danielson, 1996, 2007); Classroom Assessment Scoring System (CLASS) (Pianta et al., 2008); Stallings Classroom Observation

4 Aunque el progreso estudiantil puede evaluarse por medio de pruebas específicas, el uso de medidas de objetivos de aprendizaje se fundamenta, precisamente, en la dificultad de evaluar a todos los estudiantes, en todas las materias, todos los años mediante pruebas estandarizadas.

5 De acuerdo con Steinberg y Donaldson (2016), el 52% de los estados y el 39% de los distritos más grandes en Estados Unidos utilizan este instrumento. Además, el peso relativo asignado a este instrumento varía entre el 8% y el 25% dentro del puntaje total de la evaluación.

System (Stallings y Knight, 2003); o TEACH (Molina et al. , 2018), muchas veces los sistemas deben encontrar la rúbrica más adecuada de acuerdo con los objetivos de la evaluación, el contexto en el que esta se enmarca y el tiempo y los recursos disponibles para realizarla.⁶

En ese sentido, Donaldson y Papay (2015) y Steinberg y Donaldson (2016) señalan que este diseño implica la toma de decisiones importantes como, por ejemplo, el número de observaciones que tendrá cada docente, el tiempo de duración de cada observación, si serán observaciones programadas o sorpresa, quién será el encargado de efectuarlas (un superior, un par dentro de la escuela, un par de otra escuela o un experto externo a la escuela) y el tipo de entrenamiento que se le dará a cada observador o codificador, entre otros aspectos. Estas decisiones, además, tienen consecuencias directas tanto en la validez y la confiabilidad de la información recolectada por medio de este instrumento (Cohen y Goldhaber, 2016) como en el tiempo y el dinero que requerirá su implementación. Por ejemplo, el instrumento de observación CLASS genera datos sobre la calidad de la interacción pedagógica de los profesores, el apoyo emocional a los estudiantes y la manera en la que los docentes gestionan la clase, mientras que el instrumento de Stallings captura principalmente el uso que hacen los docentes del tiempo de instrucción y algunas prácticas pedagógicas. Debido a que el instrumento CLASS busca capturar un mayor espectro del comportamiento docente, su implementación, aunque es más costosa que la del instrumento de Stallings, permitiría una mayor retroalimentación que mejore las prácticas pedagógicas de los maestros evaluados. Sin embargo, como argumentan Bruns, De Gregorio y Taut (2016), en sistemas educativos de América Latina, donde un porcentaje alto de docentes aún no alcanza los tiempos efectivos de enseñanza mínimos sugeridos por la OCDE, puede ser beneficioso comenzar con la implementación de instrumentos más sencillos.

A pesar de estas complicaciones, la utilización de la observación en clase como instrumento tiene varias ventajas. A diferencia de las medidas de valor agregado, las observaciones en clase se enfrentan a menores niveles de controversia por parte de los docentes, para quienes los resultados obtenidos a través de pruebas estandarizadas restringen los objetivos de la labor docente a una única medida y además no les dan información relevante acerca de qué mejorar en su labor diaria o cómo hacerlo. La observación en clase, por el contrario, está arraigada en la profesión y si es implementada correctamente puede aportarle al docente información concreta acerca de sus debilidades y fortalezas. Estudios como el Proyecto de Medición de la Efectividad Docente (MET, por sus siglas en inglés) o el de Rockoff y Speroni (2010), entre otros, sugieren que estas medidas son válidas —están correlacionadas con el desempeño estudiantil en pruebas estandarizadas— y confiables cuando se diseñan e implementan correctamente.

Aunque es el instrumento más antiguo y, posiblemente, uno de los más utilizados en el mundo, aún no existe suficiente evidencia rigurosa sobre sus resultados (Cohen y Goldhaber, 2016). Por ejemplo, se desconoce si sus resultados también están relacionados con impactos

⁶ Las observaciones en aula más efectivas son aquellas que se enfocan en lo que hace efectivamente el docente en el aula, es decir, en la calidad de los procesos que ocurren en ella y menos en aspectos de infraestructura.

de largo plazo en los estudiantes, algo que, se sabe con certeza, ocurre con las medidas de valor agregado (Chetty et al., 2014b). Además, se requiere mayor investigación acerca de cuáles son las prácticas cruciales que se deben medir y por lo tanto fortalecer mediante programas de formación. Del mismo modo, si bien las conclusiones del proyecto MET reseñadas más adelante son valiosas, se necesita aún mayor investigación con respecto al diseño más adecuado y costo-eficiente a ser aplicado dependiendo del objetivo específico que se quiera alcanzar.

Portafolio docente

Un portafolio docente es la compilación de materiales por parte del docente que permite entender el proceso de enseñanza que este lleva a cabo durante el año lectivo. La ventaja de esta herramienta es que abre la posibilidad de una visión holística de la labor docente al incluir componentes de planeación, evaluación y reflexión de su labor y al recolectar evidencia acerca de la calidad de su práctica y de cómo esta se adapta a las necesidades y el progreso de sus estudiantes y depende de estos elementos. Por ejemplo, en Chile el portafolio está compuesto de tres módulos: i) descripción de la práctica docente y reflexión acerca de su quehacer docente, ii) una clase grabada y iii) descripción del trabajo colaborativo entre docentes y reflexión al respecto.

Se trata de un instrumento que, al parecer, resulta particularmente útil para la autorreflexión. Sin embargo, son pocas las investigaciones que han evaluado su efectividad. Una excepción es el estudio para Estados Unidos de Goldhaber y Anthony (2007), quienes evalúan el impacto del National Board for Professional Teaching Standards (NBPTS), el sistema de certificación docente voluntario más reconocido en el país, que se basa en la elaboración de un portafolio de evidencias de enseñanza. Los resultados alcanzados a partir de datos del estado de Carolina del Norte sugieren que el sistema es capaz de identificar a docentes más efectivos tanto en el conjunto de quienes aplican al programa como entre los que nunca lo hacen. También existe evidencia de las ventajas del portafolio para América Latina: en el análisis acerca de la correspondencia entre las distintas herramientas de la evaluación docente en Chile, Taut et al. (2016) encuentran una correlación positiva y significativa de los resultados que obtienen los docentes en el puntaje del portafolio y medidas de valor agregado de sus estudiantes. No obstante, como el portafolio en Chile incluye una clase grabada, además de las evidencias y reflexiones sobre la práctica docente y el trabajo colaborativo, no es claro si la correlación positiva proviene principalmente de los componentes de evidencias y reflexiones o si se debe al componente de observación en clase.

Encuestas a estudiantes y a otros actores

Evidencia reciente surgida del proyecto MET y de estudios posteriores que usan sus datos (Polikoff, 2015) sugiere que las encuestas a estudiantes pueden ser un complemento importante en las evaluaciones docentes. Como ocurre con los demás instrumentos, si están bien

diseñadas, las encuestas pueden tener una ventaja en términos de confiabilidad pues dependen de información recolectada por varios actores, otorgan información desde otra perspectiva y son mucho menos costosas que los otros instrumentos. Si bien existe la posibilidad de que se enfrenten a algunos problemas de sesgo, estudios para estudiantes y docentes en educación superior indican que son un instrumento válido (Benton y Cashin, 2012). Otros de los instrumentos que utilizan los distintos sistemas educativos son las entrevistas con los directores o superiores jerárquicos, las entrevistas con pares y las encuestas a padres, entre otros.

Pruebas estandarizadas a los docentes

Las pruebas estandarizadas aplicadas a los propios docentes son una herramienta útil que, dependiendo de su diseño, permite identificar su conocimiento pedagógico o disciplinar. Como quedará claro en los capítulos posteriores, el uso de este tipo de instrumento es común dentro de los sistemas educativos de América Latina y el Caribe, particularmente en las evaluaciones asociadas al ingreso a la carrera y en aquellas requeridas para ascender. Es un instrumento común también en los sistemas educativos dentro de Estados Unidos donde es un requisito para la certificación docente. Probablemente, la mayor ventaja de este instrumento es que, una vez su diseño asegure que las preguntas incluidas efectivamente evalúan si el docente cumple o no con los estándares de desempeño requeridos, su implementación es mucho más sencilla y menos costosa que la de los otros instrumentos arriba referenciados. Además, es el único instrumento para el cual hay evidencia causal positiva de la región que muestra cómo su implementación puede efectivamente mejorar el desempeño estudiantil (Ome, 2012; Brutti y Sánchez, 2017; Estrada, 2019).

Estudios que analizan diferentes instrumentos

La abundancia de instrumentos existentes presenta a las autoridades educativas una clara disyuntiva al momento de elegir cuántos y cuáles de ellos incluir dentro de sus sistemas de evaluación docente. Además, luego de dar respuesta a estas dos preguntas los formuladores de políticas deben decidir qué peso asignarle a cada uno de ellos dentro de la calificación final. Estas decisiones tendrán impactos en la validez y la confiabilidad de las evaluaciones, por eso deberían basarse en la evidencia disponible en cada contexto y, en ausencia de ello, en la evidencia internacional.

Uno de los estudios más influyentes y extensos sobre el tema es el del proyecto MET, que nació con el objetivo de construir y probar medidas de efectividad docente que sirvieran de insumos para ser utilizados en procesos de evaluación docente, que retroalimentaran las prácticas de los profesores en el aula y aumentarían su eficacia.⁷ El proyecto trabajó

⁷ Los resultados se encuentran disponibles en: <http://k12education.gatesfoundation.org/blog/measure-of-effective-teaching-project-faqs/>.

con cerca de 3.000 docentes de siete distritos escolares en Estados Unidos entre 2009 y 2013. Se analizaron tres tipos de instrumentos: observaciones de clases (con protocolos de observación tanto genéricos como específicos para ciencias y matemáticas), encuestas de percepción a estudiantes y medidas de ganancias de aprendizaje de estudiantes con base en exámenes estandarizados.⁸ Una de las conclusiones que se pueden extraer del estudio es la validez de las encuestas a estudiantes y de los protocolos de observación en clase que, cuidadosamente diseñados y aplicados, permiten identificar y dar retroalimentación a prácticas pedagógicas que a su vez están correlacionadas con el nivel de aprendizaje de los estudiantes. En ese marco, a continuación, se reseñan tres grandes preguntas que los investigadores buscaron responder en el proyecto MET (2013).

El primer cuestionamiento gira en torno de si medidas de enseñanza efectiva pueden identificar, o no, a los docentes que propician mayores aprendizajes en sus estudiantes. Para averiguarlo, los analistas recogieron medidas de eficiencia de docentes con base en observaciones en clase, encuestas a estudiantes y medidas de valor agregado ajustadas por las características socioeconómicas de los estudiantes para el período académico 2009-10 y construyeron una medida compuesta. En el siguiente año llevaron a cabo un experimento donde asignaron estudiantes a cada profesor de manera aleatoria para después determinar si aquellos estudiantes asignados a docentes identificados como más efectivos de acuerdo con la medida compuesta aprendían más. Los resultados sugieren que efectivamente estos estudiantes tuvieron un desempeño mejor en pruebas estandarizadas tradicionales e incluso en pruebas con mayores niveles de requerimientos analíticos que estudiantes asignados a docentes menos efectivos.

La segunda gran pregunta que buscaron responder en el proyecto MET se refiere a la ponderación óptima que cada instrumento debe tener dentro de una evaluación de desempeño. Debido a la complejidad del trabajo docente, es claro que realizar una evaluación con una multiplicidad de instrumentos de calidad es mejor que hacer una evaluación con un único instrumento. Sin embargo, no existen muchos estudios que permitan identificar qué combinación es la más efectiva. Los investigadores optaron por cuatro ponderaciones distintas que asignaban pesos alternativos a las medidas de valor agregado, a las observaciones en clase y a las encuestas a estudiantes. De acuerdo con los autores, una ponderación equitativa entre las tres medidas logra la mayor correlación con resultados en

⁸ Los protocolos de observación en clase incluyeron el Framework for Teaching de Danielson; Classroom Assessment Scoring System (CLASS), de la Universidad de Virginia; Mathematical Quality of Instruction (MQI), de la Universidad de Michigan; Protocol for Language Arts Teaching Observations (PLATO), de la Universidad de Stanford. Además, se utilizaron protocolos específicos de ciertas áreas de conocimiento, como el Quality of Science Teaching (QST), de la Universidad de Stanford, y el UTeach Teacher Observation Protocol (UTOP), de la Universidad de Texas-Austin, para matemáticas y ciencias. Las encuestas a estudiantes fueron las Tripod Student Perception Survey, desarrolladas por Ron Ferguson. Los exámenes de conocimiento incluyeron los estatales y el Stanford 9 Open-Ended Reading Assessment, el Balanced Assessment of Mathematics (BAM) y los ACT QualityCare para álgebra, inglés y biología.

distintas pruebas estandarizadas a los estudiantes y una mayor confiabilidad a lo largo de los años, por lo cual consideran que es una de las mejores opciones.⁹ Una ponderación que asigna el 50% a observación en clase y el 50% restante, y divide equitativamente la cifra entre encuestas a estudiantes y medidas de valor agregado, otorga resultados similares. Así, el estudio MET concluyó que una combinación de instrumentos como la observación en clase, medidas de valor agregado docente y encuestas a estudiantes son válidas.¹⁰

Por último, la investigación del proyecto MET se propuso determinar cuál es la mejor manera de implementar instrumentos de observación en clase para asegurar que sean confiables, de forma que los resultados reflejen de modo consistente aspectos relacionados con las prácticas docentes y no estén sujetos a otros factores como, por ejemplo, quién es el observador o si es un día particularmente malo para el docente. La investigación permitió concluir que incrementar de una observación a dos observaciones realizadas por la misma persona aumenta la confiabilidad del instrumento de 0,51 a 0,58 (en una escala de 0 a 1). Si esa segunda observación la lleva a cabo una segunda persona, la confiabilidad aumenta aún más (a 0,67). A partir de este resultado, el proyecto MET indagó si era necesario que esta segunda observación fuese efectuada durante una clase completa (de 45 minutos) y encontró que si la observación de la segunda clase la realizan tres personas distintas en bloques de 15 minutos se obtiene el mismo nivel de confiabilidad que si la observación de la clase completa la realiza una persona.

Si bien los impactos que reflejan el proyecto MET y los estudios particulares referenciados antes son positivos, es importante resaltar casos en los que a pesar de los esfuerzos en tiempo y dinero que realizan los sistemas educativos para diseñar e implementar procesos de evaluación docente, estos pueden no generar los impactos esperados. Un caso particular lo constituye el proyecto de la Corporación RAND y el American Institute for Research, financiado también por la Fundación Gates. Este proyecto buscaba mejorar los aprendizajes de los estudiantes y su ingreso a educación superior, particularmente de aquellos jóvenes pertenecientes a grupos minoritarios, a través del aumento en el acceso a docentes excelentes. Entre 2009 y 2016, tres distritos escolares y cuatro organizaciones que administran colegios en concesión (*charter management organizations*) acordaron participar en el estudio. Todos se comprometieron a desarrollar en sus sistemas un proceso de evaluación docente detallado y a plasmar los resultados obtenidos en la evaluación en el desarrollo profesional y la carrera docente de sus profesores.

⁹ Como era de esperarse, los autores encontraron que una ponderación que asigna el 80% del peso a medidas de valor agregado basadas en pruebas estandarizadas estatales es la que mejor predice el desempeño futuro de los estudiantes en estas mismas pruebas. Sin embargo, no necesariamente es la mejor opción para capturar otros aspectos de la docencia. De hecho, es la combinación que obtiene la menor correlación con otro tipo de pruebas estandarizadas aplicadas a estudiantes.

¹⁰ Por otra parte, el estudio encontró que el puntaje obtenido por los docentes en exámenes de conocimiento específico, que incluían temas disciplinarios y pedagógicos, no estaba correlacionado con los aprendizajes de los estudiantes.

El informe final del proyecto contiene valiosos aprendizajes para todos los interesados en el tema (Stecher et al., 2018).¹¹ El primero de ellos se relaciona con la implementación de estos sistemas de evaluación. Aunque en todas las escuelas participantes se diseñaron y aplicaron sistemas de evaluación docente basados en rúbricas de observación en clase y, donde era posible, en resultados académicos de los estudiantes, su implementación no fue sencilla. Por ejemplo, en la mayoría de las escuelas las observaciones estaban a cargo de los directores, lo cual aumentó de manera importante su carga laboral. Esta situación llevó a que en algunos sistemas se bajara el número o la duración de las observaciones o bien se designara como agente evaluador otro actor interno sin el mismo nivel de preparación para llevar a cabo dicha labor. Además, como sucede en muchos de los sistemas que se analizan en los siguientes capítulos, la mayoría de los docentes fue catalogada en las categorías de desempeño más eficientes, hecho que disminuyó el poder de diferenciación de dicho instrumento en cuanto a la calidad de los docentes.

De manera consistente con otros estudios, el proyecto encontró que la implementación de la evaluación docente aumentó la tasa de salida del sistema de los docentes poco efectivos. Sin embargo, no logró la retención o la contratación de docentes más efectivos. Esto puede explicarse por el hecho de que el uso de los resultados no fue el esperado. Por ejemplo, aunque se implementaron algunas políticas que asociaban los resultados a ascensos de carrera o bonificaciones estas no se dieron en la extensión que se esperaba, su diseño difería de lo sugerido por los investigadores al inicio del proyecto y en algunos casos estaban asociadas a aspectos que no eran importantes para los docentes.¹² Por otra parte, los sistemas buscaron integrar programas de desarrollo profesional basados en los resultados, pero la investigación mostró que no supieron cómo llevarlos a cabo y su implementación no fue completa. Según explicaron las autoridades de los establecimientos que participaron del proyecto, los resultados y la retroalimentación de estos no estaban suficientemente detallados para diseñar programas de desarrollo profesional adecuados, no existían programas que sirvieran como modelos y los sistemas educativos no tuvieron el tiempo ni los recursos monetarios para diseñarlos. Además, cuando los programas generales de desarrollo profesional y formación en servicio existían los docentes no tenían obligación de asistir, no había supervisión acerca de quiénes los tomaban y tampoco existía un control sobre si las prácticas pedagógicas de los docentes que participaban en los programas mejoraban como resultado de ello.

Teniendo en cuenta estos resultados, no sorprende que la investigación no encontrara impactos en el acceso a docentes más efectivos, en el desempeño de estudiantes en

¹¹ Los resultados del estudio se encuentran disponibles en: <https://www.rand.org/education-and-labor/projects/evaluating-teaching-effectiveness.html>.

¹² Por ejemplo, el estudio menciona que las bonificaciones por la obtención de resultados excepcionales no eran significativas o que si bien se crearon algunos roles específicos para docentes con resultados sobresalientes en las evaluaciones no se implementó un sistema de ascensos en la carrera docente como lo imaginaban los investigadores al comienzo del proyecto.

pruebas estandarizadas y en las tasas de graduación. No obstante, según los investigadores, el análisis de estos resultados debe contemplar que en tres de los cuatro estados donde se realizó el estudio se implementaron sistemas de evaluación con consecuencias de manera generalizada, por lo cual los grupos de control se vieron contaminados. Además, los autores sugieren que puede no haber pasado el tiempo necesario para observar resultados. De todos modos, el estudio deja enseñanzas importantes para los formuladores de políticas y académicos interesados en el tema: i) si la implementación de sistemas de evaluación no está acompañada de mecanismos fuertes para hacer un uso efectivo de los resultados queda básicamente incompleta; ii) como se discutió anteriormente, los autores resaltan que los sistemas reportaron dificultades al tratar de utilizar los mismos resultados para alcanzar tanto el objetivo del desarrollo profesional como el de las implicaciones en la carrera profesional.

EVALUACIONES DE IMPACTO DE SISTEMAS DE EVALUACIÓN DOCENTE

Las evaluaciones de impacto pretenden responder a una pregunta específica de causa y efecto: ¿cuál es el impacto (o efecto causal) de una práctica o de un programa en un resultado de interés? (Getler et al., 2016). A pesar de que todos los países evalúan, de una manera o de otra, a sus docentes, el conocimiento acerca de los efectos causales que estas prácticas pueden producir en su desempeño laboral y, en última instancia, sobre el aprendizaje de sus alumnos, es escaso. Como se ha mencionado antes, la evaluación docente y el uso adecuado de los resultados por medio de programas de formación e incentivos laborales buscan mejorar el aprendizaje de los alumnos. Esta sección sintetiza las principales conclusiones que se pueden extraer de las evaluaciones de impacto que existen hasta el momento, sobre la base de la revisión de los tres pasos de la teoría de cambio del gráfico 2.1. Primero se resumen los impactos en la fuerza docente a través de los cambios compositionales y la mejora en las prácticas docentes, y después se mencionan los impactos en el aprendizaje estudiantil.¹³

Impactos en la composición del cuerpo docente, en la retención de los docentes y en las habilidades pedagógicas de los profesores

Dos estudios para Estados Unidos sugieren que sistemas de evaluación docente holísticos pueden tener impactos importantes en la composición y la retención de la fuerza docente.

¹³ Resulta interesante también analizar la literatura que evalúa el impacto de los incentivos monetarios individuales o grupales que reciben los profesores a partir del desempeño de sus estudiantes. Como no se centra en evaluar las destrezas pedagógicas y disciplinares de los docentes, ese análisis no se incluye en este trabajo. Sin embargo, vale mencionar estudios como Springer (2009), Fryer (2013), Bruns y Luque (2015) y Barrera-Osorio y Raju (2017).

El primer estudio es el de Dee y Wyckoff (2015), quienes analizan el impacto que tuvo el sistema de incentivos atados a la evaluación docente (conocido como IMPACT) en las escuelas públicas del distrito de Columbia (EPDC) en la rotación de docentes catalogados como efectivos y no efectivos por el mismo sistema y en el desempeño de los docentes que permanecen después de haber sido evaluados. El sistema de evaluación IMPACT fue introducido de manera general en el año académico 2009-10 con los objetivos explícitos de mejorar la calidad de la enseñanza y el aprendizaje de los alumnos. En el marco de IMPACT, los docentes recibían una calificación final que dependía en gran medida de resultados obtenidos a partir de observaciones en clase rigurosas atadas a los estándares de desempeño del distrito (*teaching and learning framework*). La calificación iba de 100 a 400 puntos y el puntaje obtenido definía la ubicación de cada docente en una de cuatro posibles categorías de desempeño: no efectivo (100-175), mínimamente efectivo (176-250), efectivo (251-350) y altamente efectivo (351-400).

El sistema contemplaba cinco observaciones en el año: tres las realizaba un directivo de la escuela (director o asistente del director) y dos las efectuaba un experto externo. La única observación anunciada era la primera, que la llevaba a cabo el director de la escuela. Para un grupo reducido de docentes (el 17%) cuyos estudiantes presentaban pruebas estandarizadas también se incluían medidas de valor agregado basado en esas pruebas. Para los que no tenían la posibilidad de cálculo de valor agregado se incorporaba una medida de objetivos de aprendizaje para los estudiantes fijados por el docente y el director. Al finalizar el año los administradores de la escuela evaluaban qué tanto se habían alcanzado esos objetivos. Además, existían dos medidas adicionales: el compromiso con el colegio y la comunidad y una medida de valor agregado para el colegio. Finalmente, el diseño incluyó mecanismos de mejoras por medio de tutores para los docentes que, de acuerdo con los resultados obtenidos, así lo requirieran.

El sistema comprendía dos tipos de incentivos basados en los resultados que cada docente obtenía en la evaluación. El primero consistía en un incentivo monetario con una bonificación para aquellos docentes clasificados como altamente efectivos. Asimismo, el sistema aseguraba un aumento permanente en la base salarial (equivalente a entre tres y cinco años de servicio) para los docentes catalogados como altamente efectivos en dos oportunidades seguidas y retiraba de manera inmediata a los docentes clasificados como no efectivos. El segundo tipo de incentivo estaba relacionado con la estabilidad laboral. En ese marco, los docentes clasificados en la categoría de mínimamente efectivo por dos años consecutivos eran removidos del sistema también. Así, ambos incentivos de IMPACT no solo estaban ligados a medidas multidimensionales de desempeño académico asociadas a mayores niveles de confiabilidad y transparencia —como se explicó anteriormente— sino que además se basaban en acciones que los propios docentes controlan de forma directa.

Para estimar el impacto que tuvo la evaluación los investigadores compararon la permanencia y el desempeño de docentes cuyos puntajes en la evaluación del año anterior estuviesen cerca de los límites que separaban a un docente efectivo de uno mínimamente

efectivo (alrededor de 250 puntos) y en el límite entre los docentes clasificados como efectivos y los altamente efectivos (350 puntos).¹⁴ Los resultados demuestran que la amenaza de despido aumentó el retiro voluntario de docentes clasificados como mínimamente efectivos en la primera evaluación. Además, el desempeño de los docentes clasificados en esta categoría que permanecieron en el sistema mejoró en la evaluación siguiente, medido por el puntaje de la evaluación en años posteriores. Por último, los autores encuentran que los incentivos monetarios entregados a aquellos ubicados en la cola superior de la distribución fueron efectivos al lograr mejorar sus resultados en las siguientes evaluaciones también. Por otra parte, el estudio revela que los impactos crecen en el tiempo, a medida que las bonificaciones y las amenazas se vuelven más creíbles. Sin embargo, los autores no pueden diferenciar qué parte de los impactos provino exclusivamente del uso de la evaluación y qué parte de los incentivos.

Sartain y Steinberg (2016) responden a una pregunta similar a partir del análisis de los impactos de un sistema de evaluación docente piloto, llamado Excellence in Teaching Project (EITP). Este programa, aplicado en un grupo de escuelas públicas de Chicago, tenía como objetivo mejorar la calidad de aprendizaje de los estudiantes a través de una mejora en la retroalimentación brindada por los directores a sus profesores para de esa manera mejorar las prácticas docentes en clase. El programa utilizó la rúbrica de Danielson (Framework for Teaching, FFT) e implementó un entrenamiento profundo que capacitaba a los directores en el uso correcto de la misma. Por otra parte, los docentes no estaban sujetos a un beneficio o un castigo específicos basados en los resultados obtenidos. Los autores analizan el impacto del sistema de evaluación en dos aspectos: i) la salida de los docentes de la escuela en la que trabajan, pero no del sistema de escuelas públicas de Chicago (rotación dentro del distrito) y ii) la salida definitiva de los docentes del sistema de escuelas públicas de esa ciudad.

En este caso, para encontrar los impactos se llevó a cabo un experimento aleatorio por el cual se eligieron 92 escuelas públicas: 44 de esas escuelas recibieron el tratamiento en el período académico 2008-09 y las 48 escuelas restantes, en el año académico siguiente. Por lo tanto, se trata de una evaluación de muy corto plazo en la que los indicadores de interés se compararon un año después de que el programa fue implementado en las primeras 44 escuelas. Los resultados sugieren que el EITP no tuvo ningún impacto en la tasa promedio de rotación de docentes dentro del distrito. Sin embargo, al separar los docentes de acuerdo con su nivel de eficacia (según evaluaciones previas a 2009), los autores encuentran que la probabilidad de salida del sistema de escuelas públicas de Chicago aumenta para aquellos clasificados con nivel no satisfactorio o satisfactorio, resultado muy similar al hallado recientemente por Stecher et al. (2018). También determinan que la probabilidad de salida del sistema es mayor para aquellos docentes con pocos años de experiencia. Así, los autores concluyen que un sistema de evaluación como este tiene el potencial de retirar del sistema educativo a

¹⁴ Esta es una metodología cuasi-experimental, ampliamente utilizada en la literatura especializada, conocida como la metodología de Regresión Discontinua.

aquellos docentes con bajos niveles de eficacia. Por otra parte, los autores señalan que los nuevos docentes contratados en las escuelas tratadas, al igual que aquellos que permanecieron en ellas después de implementado el sistema de observación y retroalimentación de prácticas pedagógicas, tienen un mayor puntaje en el FFT que los docentes que se fueron.

Ambos estudios, por lo tanto, sugieren que observaciones de clase con rúbricas atadas a estándares claros, observadores cuidadosos e intensamente entrenados y una retroalimentación continua para los docentes de sus prácticas pueden lograr mejoras en la calidad de la enseñanza y retirar de las aulas a los docentes con prácticas deficientes. Más aún, para el caso de Chicago, el entrenamiento de los propios directores en cómo evaluar a sus docentes produjo impactos positivos en la calidad de los docentes contratados en el sistema. Estos resultados, además, son congruentes con evidencia previa disponible también para Estados Unidos en la que se observa que la entrega de información a los directores de escuela acerca de la efectividad relativa de los docentes, medida a partir de estimativos de valor agregado, impacta en la retención de docentes poco efectivos y, a través de este canal, en los resultados de los estudiantes (Rockoff et al., 2012).

Impactos en los resultados académicos de los estudiantes

Como se puede apreciar en el gráfico 2.1, el objetivo final de la evaluación docente es asegurar que el sistema cuente con docentes excelentes que promuevan de manera efectiva el aprendizaje de los alumnos. El estudio de Taylor y Tyler (2012) analiza esta posibilidad con datos de las evaluaciones docentes realizadas en Cincinnati, Estados Unidos. Los autores aprovechan la aplicación progresiva de las evaluaciones para comparar un grupo de docentes que participó del programa primero con otro grupo que empezó a participar más adelante. El programa de evaluación del estado contemplaba cuatro observaciones de clases metódicas durante un año académico. Tres de ellas las realizaba un par experto, externo al colegio —escogido según sus propios niveles de eficacia y designado para tal fin de manera expresa—, y la cuarta la efectuaba el director u otro directivo de la misma escuela. Las observaciones se basaban en la rúbrica de Danielson; en tanto que todos los observadores contaban con un entrenamiento riguroso acerca de cómo llevarlas a cabo. La primera observación se anunciaba durante la semana en la que ocurriría, pero las siguientes observaciones no eran informadas con anticipación. Al terminar el proceso anual cada docente recibía una calificación, en el marco de un reporte de desempeño final, en uno de cuatro niveles: distinguido, eficaz, básico y no satisfactorio. Los resultados de las evaluaciones tenían pocas implicaciones en la estabilidad de la carrera: niveles excelentes podían estar atados a promociones, niveles bajos originaban la asignación de pares para un programa de mejoramiento, y en muy pocos casos generaban la terminación del contrato laboral. Además, cerca del 90% de los docentes fue ubicado en las dos primeras categorías, lo que ratifica el escaso impacto en términos de carrera que las evaluaciones per se podían traer.

Los autores utilizan una metodología que les permite analizar cómo cambia la efectividad de cada docente antes de pasar por el proceso de evaluación, durante el mismo y después de haberlo atravesado.¹⁵ Así, encuentran que la productividad de los docentes, medida como el aprendizaje basado en el puntaje de un estudiante de su clase en el área de matemáticas, aumenta durante el año en el que son evaluados e incluso se incrementa en mayor proporción en los años posteriores a la culminación del proceso de evaluación. Un estudiante sube su puntaje en pruebas estandarizadas de matemáticas en 0,1 desviaciones estándar más que un estudiante con características socioeconómicas similares y niveles de conocimiento previos similares que fue asignado al mismo profesor, pero antes de que este pasara por el proceso de evaluación. Aunque los datos no permiten identificar los canales a través de los cuales se pueden explicar los impactos, los autores sugieren que se basan en el aprendizaje que los docentes obtienen gracias al proceso evaluativo acerca de su rendimiento y sus prácticas pedagógicas, y el subsecuente desarrollo de habilidades adicionales, así como el mayor esfuerzo futuro.¹⁶

Steinberg y Sartain (2015) analizan el impacto del sistema de evaluación docente piloto EITP en Chicago (que ha sido descrito más arriba) sobre el aprendizaje de los estudiantes. Como se ha mencionado antes, el piloto comenzó en 2008-09 en 44 escuelas primarias (primera cohorte) y en 2009-10 se incluyeron 48 escuelas adicionales (segunda cohorte). Los autores encuentran que los estudiantes que asistían a los colegios de la primera cohorte obtienen mejores resultados en lectura y matemáticas que aquellos estudiantes que asistían a colegios cuyos docentes fueron tratados en la segunda cohorte, aunque únicamente los resultados en lenguaje son estadísticamente significativos. Además, hallan que dicha ventaja se mantiene incluso después de haber sido tratados los de la segunda cohorte y que los impactos fueron mayores para escuelas donde asisten estudiantes con mayores niveles de ingreso promedio.

En sus conclusiones los autores resaltan tres puntos importantes. El primero es que se pudieron obtener resultados inmediatos en el aprendizaje estudiantil, gracias a que el sistema estaba basado en observaciones de clase realizadas por los directores de las escuelas, quienes después discutían de manera detallada los resultados y las posibles estrategias de mejoramiento con los docentes. Sin embargo, este esquema tuvo dos implicaciones: i) el entrenamiento de los directores, y en general de la persona encargada de la observación y la retroalimentación, debe ser intenso y de la mejor calidad; ii) la implementación de la evaluación y las discusiones posteriores generó en los directores una presión muy importante

15 En términos técnicos, los autores utilizan una estrategia de efectos fijos por maestro. Además, en sus estimaciones tienen en cuenta los puntajes previos de sus alumnos, puntajes que no entraban en el proceso de evaluación del estado de Cincinnati que estaban analizando.

16 Este mismo análisis se llevó a cabo con docentes en el área de lengua a través de pruebas de lectura. En este caso los autores no encuentran ningún impacto positivo y no ofrecen explicaciones al respecto, al margen de sugerir que los estudiantes aprenden y practican la lectura en contextos que están fuera de la escuela mucho más que matemáticas y que, por lo tanto, la eficacia o las diferencias en docentes en esta área de la enseñanza son de por sí menores.

en tiempo y trabajo que los formuladores de políticas deben tener en cuenta. Finalmente, los autores señalan que, en general, la implementación de la evaluación se debilitó de manera significativa a partir del segundo año, hasta el punto de que el programa nunca fue aplicado de manera general luego del piloto. Esto, de acuerdo con ellos, explica los menores impactos observados para la cohorte dos y pone de manifiesto la necesidad de contar con un sistema fuerte para que sea posible obtener resultados positivos. En ese sentido, resaltan la importancia de una implementación adecuada y fiel al diseño original.

Adnot et al. (2017) también hallan resultados positivos en el aprendizaje de los estudiantes al analizar el impacto del retiro de profesores poco eficientes incentivado por el programa IMPACT.¹⁷ Los autores encuentran que el retiro de docentes en el distrito mejoró en 0,08 desviaciones estándar y 0,04 desviaciones estándar el desempeño de los estudiantes en matemáticas e inglés, respectivamente. Incluso si solo se tiene en cuenta el retiro de los docentes clasificados como poco efectivos en la evaluación, las mejoras promedio en los resultados académicos de los estudiantes aumentan a un 21% y un 14% en cada materia, respectivamente. Además, los autores mencionan que la recomposición en la efectividad del grupo de docentes en el sistema fue mayor en escuelas con altos niveles de población vulnerable, por lo cual las mejoras en el aprendizaje se concentran en estos estudiantes. A su vez, esto implica que los impactos en este tipo de alumnos son aún mayores y equivalen a cerca de un 35%–65% de su aprendizaje promedio en un año.

Para América Latina la evidencia acerca de los impactos que ha generado la evaluación docente es mucho menor y se concentra en casos particulares. Sin embargo, parte de ella sugiere que las evaluaciones docentes pueden provocar impactos positivos sobre los aprendizajes de los estudiantes. Por ejemplo, Ome (2012) y Brutti y Sánchez (2017) en Colombia y Estrada (2019) en México encuentran que evaluaciones en la entrada a la carrera docente tienen impactos marginales positivos en indicadores de eficiencia y calidad de la educación que reciben los estudiantes. Para Chile, aunque no es posible catalogarlo como una evaluación de impacto del sistema de evaluación docente, el estudio de Taut et al. (2016) halla evidencia que sugiere que los aprendizajes logrados por los estudiantes cuyos docentes obtuvieron puntajes más altos en el instrumento del portafolio son mayores que los aprendizajes alcanzados por los estudiantes asignados a otros docentes.

Aunque no analiza un sistema de evaluación docente como tal, un estudio relacionado es el de Santibáñez et al. (2006), pues estima el impacto que generó en México la introducción del programa Carrera Magisterial. Este programa voluntario de incentivos salariales, implementado en el país entre 1992 y 2015, daba la posibilidad a los maestros de alcanzar una promoción o ascenso horizontal. Las promociones horizontales son aquellas en las que el prestigio, las funciones, las responsabilidades y el salario de los docentes pueden

¹⁷ Los autores utilizan información a nivel de grado-escuela y emplean una metodología de diferencias en diferencias que evalúa el impacto que el retiro de profesores de ese grado-escuela particular, incluidos aquellos que salieron del sistema debido a IMPACT, tiene en el rendimiento estudiantil.

incrementarse sin necesidad de que los profesores deban abandonar las aulas de clase o el enfoque en la docencia. Para poder acceder a un ascenso, los maestros interesados eran evaluados anualmente en seis dimensiones: i) grado máximo de estudios, ii) años de antigüedad en el servicio, iii) desempeño profesional de acuerdo con la calificación otorgada por sus pares, iv) puntaje en las evaluaciones estatales y federales de los cursos de desarrollo profesional, v) puntaje en las pruebas de conocimiento de docentes y vi) puntaje promedio en pruebas estandarizadas de sus alumnos. Según la clasificación final obtenida, los maestros eran catalogados en seis posibles rangos que a su vez determinaban el monto de las bonificaciones salariales a las que tenían derecho.

Los autores analizan tanto la confiabilidad y la consistencia interna de las pruebas aplicadas como los impactos que el programa provocó en la calidad educativa que recibieron los alumnos. En ese sentido señalan que, si bien las pruebas de maestros estaban alineadas con el currículo y presentaban grados de confiabilidad y consistencia interna adecuados, en realidad el nivel de complejidad de las preguntas era bajo y había espacio para mejorar. Respecto de los exámenes a estudiantes, los autores mencionan que la confiabilidad y la consistencia interna eran menores que los observados en las pruebas docentes, y la habilidad cognitiva exigida también era escasa. Teniendo en cuenta esto, los impactos de Carrera Magisterial en aprendizaje estudiantil son mixtos. Aunque en secundaria se hallan mejoras de entre el 3% y el 15% de una desviación estándar en las pruebas estandarizadas de alumnos cuyos maestros fueron clasificados en el grupo de mayores incentivos, no se encuentra ningún impacto para estudiantes de primaria. Incluso se advierten algunos impactos negativos para alumnos de primaria cuyos maestros ya habían sido clasificados en el grupo de mayores incentivos en años anteriores.

En su conjunto, la evidencia acerca de las evaluaciones docentes indica que tienen el potencial de mejorar la calidad educativa en los sistemas en los que se aplican. Sin embargo, para que esto efectivamente ocurra es esencial que los formuladores de políticas tengan en cuenta que el diseño y la implementación son factores fundamentales. Si el diseño y la implementación de las evaluaciones docentes no aseguran que los instrumentos y la información que surja de ellas sean válidos y confiables, es poco probable que los recursos invertidos en ellas tengan frutos verdaderos. Incluso si esta primera condición se cumple es importante también el uso que se les dé a los resultados que arrojen las evaluaciones. Aunque hay evidencia que sugiere que la efectividad docente mejora a través de la participación en procesos evaluativos, la teoría de cambio indica con claridad que es necesario incluir procesos formativos o incentivos en la carrera ligados a las evaluaciones. Esto implica que las decisiones que es preciso tomar alrededor del tema no solo son complejas sino también costosas. Entender cómo han enfrentado esta disyuntiva otros sistemas educativos puede ser informativo y enriquecedor para la toma de decisiones futuras. Esto es precisamente lo que se busca en los siguientes dos capítulos, en los que se analiza de qué manera sistemas educativos de países que se considera que gozan de un buen desempeño educativo y 19 países de América Latina han enfrentado estos dilemas de política pública educativa y docente.



CAPÍTULO 3

CÓMO SE EVALÚA A LOS DOCENTES EN LOS SISTEMAS EDUCATIVOS EXITOSOS

El capítulo anterior dejó plasmada la multiplicidad de decisiones que deben tomar los formuladores de políticas al momento de diseñar, implementar y utilizar la información de los sistemas de evaluación docente. Esta multiplicidad sugiere además que no puede ni debe existir una única solución a estas disyuntivas para todos los sistemas educativos. Cada país está inmerso en contextos distintos y los sistemas de evaluación docente finalmente implementados deben concordar con los objetivos, prioridades y posibilidades que cada gobierno quiera y pueda alcanzar con ellos. Lo que funciona en países de buen desempeño educativo no necesariamente va a funcionar en los sistemas educativos de América Latina y el Caribe, ya sea por razones presupuestales, por la experiencia en la realización de evaluaciones, o por las particularidades del cuerpo docente, entre otras causas. Sin embargo, el análisis de las principales características de los sistemas de evaluación docente de países que son reconocidos por su calidad educativa puede brindar ideas valiosas respecto de su diseño, implementación y de las consecuencias que de allí emanan.

Eso es lo que precisamente hace este capítulo, al revisar de manera general la historia reciente y las principales características de las evaluaciones docentes en países que se considera que gozan de un buen desempeño educativo.¹ Específicamente se analizan los sistemas de evaluación docente para Ontario (Canadá), Reino Unido, Nueva Zelanda,

¹ Este análisis se basa en gran parte en los documentos de antecedentes producidos por Wyckoff y Katz (2018) y CUREE (2018) para el proyecto “Midiendo la efectividad docente”, y en el estudio reciente de Steinberg y Donaldson (2016), quienes resumen los sistemas de evaluación docente de 50 estados y 25 distritos escolares de Estados Unidos.

Singapur, República de Corea y Estonia. Se seleccionaron estos países por su alto desempeño en pruebas estandarizadas para estudiantes y/o porque cuentan con diseños novedosos de evaluación docente. En Estados Unidos se analizan los sistemas de Denver Public Schools (DPS), District of Columbia Public Schools (DCPS), New Haven Public Schools (NHPS) y New York City Public Schools (NYCPS) y los estados de Nuevo México y Tennessee. Estos se eligieron porque incorporan medidas de valor agregado en sus evaluaciones docentes.

Para entender las decisiones finales en torno al diseño y a la implementación de los sistemas de evaluación docente, es importante comenzar por conocer la historia que se encuentra detrás de cada uno de los casos analizados. En el caso de Estados Unidos, el desarrollo de los sistemas de evaluación de desempeño docente, que utilizan una combinación de instrumentos y que buscan entender y diferenciar las prácticas pedagógicas en las escuelas, es un fenómeno relativamente reciente. De acuerdo con Wyckoff y Katz (2018), para impulsar la introducción de estos sistemas influyó una combinación de al menos tres factores nacionales e internacionales. En primer lugar, la evidencia de pruebas de conocimiento estandarizadas, nacionales e internacionales, existentes desde los años ochenta en adelante demostró que la calidad de la educación que brindaba el sistema no era la ideal y que era preciso poner en marcha acciones concretas para mejorarla. En segundo lugar, algunos estudios del área de educación demostraban que: i) los docentes eran el insumo escolar más importante de todos para el desempeño académico de los estudiantes (Aaronson, Barrow y Sander, 2007); ii) la eficiencia/calidad de los docentes en el país variaba de manera importante incluso a nivel de la escuela (Aaronson, Barrow y Sander, 2007; Rivkin, Hanushek y Kain, 2005; Rockoff, 2004), y iii) la evaluación docente en las escuelas del país era simple, calificaba a la mayoría del cuerpo docente con un desempeño satisfactorio y no aportaba retroalimentación valiosa que permitiera mejorar la calidad de la enseñanza (Weisberg et al., 2009). En tercer lugar, la clara evidencia brindada por los dos puntos anteriores creó el espacio para que se promovieran y financiaran programas específicos con el objetivo de mejorar la calidad educativa de las escuelas y el desarrollo de sistemas de evaluación docente detallados.²

El flujo de nuevos fondos federales desde comienzos del milenio, unido a la presión de diversos sectores, influyó para que muchos estados y distritos escolares de Estados Unidos reformaran los sistemas de evaluación docente de manera importante. Además de incluir múltiples instrumentos que permitieran identificar el desempeño de los docentes, en muchos casos ataron también los resultados a consecuencias en términos monetarios o laborales. La idea, por supuesto, era basar el diseño y la implementación de estos sistemas en la evidencia disponible, que sentaba las bases para asegurar sistemas de evaluación

² Tres programas conocidos que surgieron en respuesta a la evidencia académica y el sentido de urgencia con respecto a la necesidad de mejorar la calidad educativa en el país incluyen Race to the Top (2009), el Teacher Incentive Fund (2006) y el No Child Left Behind (2001).

confiables y válidos que verdaderamente condujeran a una mejora de la calidad educativa. Sin embargo, el corto plazo que dieron los programas federales para implementar los nuevos sistemas, el elevado costo monetario que estos sistemas conllevan y la presión en contra ejercida por los sindicatos docentes debilitaron su implementación en muchos casos (Stecher et al., 2018).

La revisión de los demás sistemas educativos incluidos en el capítulo permite evidenciar también la diversidad en cuanto a la historia, la madurez y los objetivos de la evaluación docente implementada en cada uno de ellos. Por ejemplo, el sistema de República de Corea, creado en 1964 y reformado a lo largo de los años, y por última vez en 2010, determina un modelo de evaluación obligatorio para todos los docentes que lleven más de dos meses trabajando en una escuela, el cual se basa en tres componentes.³ El primer componente dentro del sistema es el de la “Evaluación de desempeño”, que comprueba las cualidades, las actitudes y el trabajo de todos los docentes de las escuelas públicas del país a través de encuestas a estudiantes y pares, y mediante la observación en clase, como se detallará más adelante. El segundo componente es el “Sistema de incentivos basado en el desempeño”, cuya función es determinar los salarios de los docentes de acuerdo con su desempeño, y buscar elevar sus niveles de motivación y la calidad educativa. Finalmente, el tercer componente es el de “Evaluación para el desarrollo profesional”, el cual se introdujo en 2010 y fue diseñado para brindar retroalimentación a los docentes y fomentar sus habilidades profesionales. Este componente de desarrollo profesional se basa en los resultados de la evaluación del primer componente, pero con el agregado de un reporte de autoevaluación detallada de los docentes, en el cual ellos mismos reflexionan acerca de sus metas, fortalezas y debilidades. Así, los tres componentes se utilizan de manera complementaria y han sido introducidos paulatinamente a medida que las autoridades educativas entendieron que un único componente no lograba alcanzar los objetivos de evaluación (detallados en la teoría de cambio del capítulo anterior, gráfico 2.1).

Aunque Singapur no cuenta con tres evaluaciones distintas, las autoridades educativas del país desarrollaron un sistema central conocido como “Plan de Desarrollo Profesional del Servicio Educativo”, que cuenta también con tres componentes complementarios: el Sistema Mejorado de Manejo del Desempeño, el Sistema de Reconocimiento a través de Bonificaciones Monetarias y un sistema de carrera con tres opciones distintas (carrera de docente de aula, carrera de líder educativo, carrera de especialista experimentado). El sistema de evaluación docente, atado al primer componente, se enunció en 2001 y finalmente se implementó en 2005. Se trata de un sistema basado en competencias que determina los conocimientos, aptitudes y habilidades que cada docente debe demostrar en cada etapa de su carrera, con un doble objetivo:

³ Así, en cierta medida es el caso más parecido a lo que ocurre en la región latinoamericana donde coexisten distintos tipos de evaluaciones vigentes en cada país cada una con distintos propósitos, tal y como quedará claro en el siguiente capítulo.

- i. Fomentar el mejoramiento profesional del docente a través de la reflexión sobre sus logros y potencial y, a través de este proceso, mejorar la calidad educativa que se ofrece a los alumnos.
- ii. Vincular el sistema de evaluación docente con las decisiones sobre el salario, las bonificaciones y la progresión en la carrera.

A diferencia de estas dos modalidades, los restantes cuatro casos analizados cuentan con sistemas de evaluación mucho menos centralizados y detallados. El sistema de evaluación docente vigente en la provincia de Ontario (Canadá) es de 2002 y su última reforma se realizó en 2010. Se trata del tercer sistema educativo en términos de un mayor nivel de centralización en las decisiones y requisitos que debe incorporar una evaluación. Sin embargo, a diferencia de los otros dos casos mencionados anteriormente, desde su inicio, la evaluación en Ontario ha buscado sobre todo un único objetivo: el desarrollo profesional a través de una evaluación diagnóstica que otorgue retroalimentación de sus prácticas de aula. En cuanto al nivel de centralización, a este le sigue Reino Unido, cuyo sistema, desarrollado en 2000 y reformado en 2006 y 2012, buscaba inicialmente asegurar la satisfacción laboral y altos niveles de excelencia en el cuerpo docente. Aunque la regulación del país detalla algunos principios básicos con los que debe contar la evaluación, en realidad la responsabilidad de esta recae en los cuerpos colegiados a nivel de cada institución.

De todos los sistemas analizados, el de Nueva Zelanda es el más reciente, y su implementación debía comenzar en 2018. Tanto su diseño como su implementación pueden presentar variaciones en las diferentes escuelas, pero el sistema debe basarse en guías especialmente ideadas por la autoridad central. Finalmente, en Estonia el sistema de evaluación que la legislación obliga a aplicar es a nivel de la escuela y no a nivel central. Esto concuerda con el hecho de que los docentes del país gozan del mayor nivel de autonomía de todos los países miembros de la Organización para la Cooperación y el Desarrollo Económicos (OCDE). A su vez, implica que poco se sabe de los criterios utilizados en las evaluaciones de las escuelas y aún menos de los docentes. Sin embargo, el gobierno emite unas guías que buscan orientar las decisiones de los directores con la inclusión de la observación de clases de sus docentes como un instrumento dentro de la evaluación interna de los colegios.

A pesar de las claras diferencias en cuanto a su historia, es posible observar tres aspectos fundamentales de los sistemas de evaluación docente para los 12 casos aquí expuestos. Específicamente, el eje del análisis de estos sistemas educativos está centrado en tres aspectos concretos: i) los instrumentos utilizados en cada uno de los casos; ii) los pesos que las autoridades educativas asignan a cada uno de dichos instrumentos dentro del puntaje final y, por último, iii) las consecuencias ligadas a los resultados de la evaluación. Esto, permitirá que el lector entienda cómo en estos casos de éxito se tomaron decisiones fundamentales al momento del diseño y de la implementación de los sistemas de evaluación docente, y además posibilitará un análisis de las principales similitudes y diferencias que existen en los sistemas de evaluación docente de la región.

INSTRUMENTOS UTILIZADOS EN LAS EVALUACIONES DOCENTES

Al profundizar en el análisis comparativo de las características de los sistemas de evaluación, vale destacar como primer aspecto que, bajo distintos diseños, los sistemas de alto desempeño estudiados ostentan como su base fundamental ciertos estándares del desempeño docente. Estos estándares describen de manera explícita lo que deben saber y deben saber hacer los docentes en el proceso de construcción del conocimiento de sus estudiantes y su quehacer diario. Por lo tanto, el diseño y la implementación de esta hoja de ruta de la profesión constituyen un paso esencial para guiar y fundamentar todos los ejes dentro de la carrera, y potenciar el crecimiento y la mejora continua de la fuerza docente. Por supuesto, los estándares son también el paso previo necesario para entender el alcance y la efectividad de los sistemas de evaluación docente.

Quizás el sistema de estándares más detallado sea el de Singapur, en cuyo caso cada etapa de la carrera está alineada con su propio conjunto de competencias debidamente especificadas y evaluadas. En Ontario, hay establecidas 16 competencias, que incluyen el compromiso del docente con el aprendizaje y el bienestar de los estudiantes, y su conocimiento profesional y su aplicación. Al igual que ocurre en Singapur, en Ontario el nivel de exigencia en cuanto al cumplimiento varía dependiendo del momento de la carrera en la que se encuentra cada docente. Por ejemplo, a los docentes principiantes se les evalúa únicamente en ocho de las 16 competencias, y su cumplimiento depende de si los docentes pueden o no avanzar al siguiente nivel.

En Nueva Zelanda existen también marcos centrales de desempeño que además varían de acuerdo con el contexto específico de algunas escuelas o grupos (por ejemplo, escuelas donde el currículo es dictado en la lengua maorí). Incluso se han desarrollado estándares de desempeño en Estonia y, aunque su uso no es obligatorio en el proceso de evaluación, estos se encuentran en las guías base ideadas por las autoridades. La definición de estos estándares y su inclusión dentro de las evaluaciones son importantes, ya que además de guiar a los propios docentes con respecto a lo que se espera de ellos, contribuyen a que los instrumentos, los actores responsables de la evaluación y los resultados que emanan sean claros, imparciales y coincidan con los objetivos de excelencia educativa únicos de cada país (OCDE, 2013).

El segundo aspecto que es posible analizar para cada uno de los 12 sistemas abarca la periodicidad de la evaluación, su obligatoriedad y quién está a cargo de su implementación. En la mayoría de los casos analizados para Estados Unidos, las evaluaciones docentes son obligatorias, anuales e implementadas por las autoridades de cada escuela. Sin embargo, en New Haven y Nuevo México para la aplicación de algunos instrumentos participan también evaluadores externos. Esta es una característica destacada que permite agregar un mayor nivel de imparcialidad y consistencia a los resultados, siempre y cuando estos actores estén correctamente entrenados para tal labor.

La variación en torno a estos dos aspectos es un poco mayor en los casos analizados en los demás sistemas de alto desempeño. Por ejemplo, en Ontario es obligatorio que

los docentes sean evaluados al menos una vez cada cinco años —aunque deben realizar un plan de aprendizaje anual—, de acuerdo con lo que decida el director de la institución educativa donde trabajen. Por su parte, tanto en Reino Unido como en Nueva Zelanda las evaluaciones docentes deben ser anuales y, acorde con lo que se explicó anteriormente, la legislación vigente les deja a los directores un espacio flexible para acomodarlas, de acuerdo con lo que consideren más conveniente para sus docentes y escuelas.

En Singapur y República de Corea las evaluaciones son anuales y su estructura está mucho más coordinada por las autoridades centrales. En Singapur, aunque los docentes son evaluados por su supervisor directo de manera regular, la evaluación formal la lleva a cabo un panel de supervisores compuesto por el director, el subdirector y los jefes de departamento, quienes los califican de manera colectiva. En República de Corea, los directores son los encargados de asegurar que las directrices impartidas por el Ministerio de Educación en torno a las evaluaciones se cumplan. Finalmente, en Estonia la evaluación interna de la institución debe llevarse a cabo cada tres años, aunque cada año debe elaborarse un reporte de progreso interno. Además, los docentes pueden aplicar de manera voluntaria a procesos que les permitan ascender en la carrera, así como también a procesos de certificación cuyas características son definidas por las escuelas y por la asociación de docentes del país, respectivamente.

Aunque la periodicidad y obligatoriedad son aspectos esenciales que permiten entender de manera general el funcionamiento de los sistemas de evaluación en cada caso, una de las decisiones fundamentales que deben tomar las autoridades educativas de cada sistema tiene que ver —como quedó claro a partir del capítulo anterior— con los tipos de instrumentos que se utilizarán. El cuadro 3.1 resume los instrumentos empleados para las evaluaciones docentes en cada uno de los 12 sistemas aquí analizados. Debido a la historia que se encuentra detrás de cada sistema, y que ha sido reseñada anteriormente, no es de extrañar que se vean reflejadas las importantes diferencias que existen en cuanto a la selección de instrumentos en cada caso.

La primera conclusión evidente del análisis de la información es que en los sistemas estudiados de Estados Unidos las evaluaciones docentes suelen incluir una multiplicidad de instrumentos, mientras que esto no ocurre en la mitad de los demás sistemas de la OCDE analizados en este capítulo. La primera excepción es Nueva Zelanda, país en donde los docentes, además de estar sujetos a la observación en clase, deben llevar a cabo también una autoevaluación que incluye un portafolio con evidencia que demuestre que el docente efectivamente está alcanzando los estándares exigidos en la profesión.⁴ Lejos de los otros cuatro sistemas, se encuentran Singapur y República de Corea, cuyas evaluaciones estructuradas abarcan el uso de al menos cuatro instrumentos. Para estos dos países, el sistema de evaluación vigente incluye, además de la observación en clase, el análisis de

4 De hecho, en el país se desarrollaron guías para la construcción de dicho portafolio.

CUADRO 3.1 | Instrumentos utilizados en las evaluaciones docentes en algunos sistemas educativos exitosos seleccionados

País	Observación en clase	Portafolio	Resultados de exámenes a estudiantes	Objetivos de aprendizaje estudiantil	Encuestas a estudiantes y/o padres	Autoevaluación	Otros
Ontario (Canadá)	■						
Reino Unido	■						
Nueva Zelanda	■					■	
Singapur	■	■	■			■	
Rep. de Corea	■				■	■	
Estonia	■						
Denver	■		■	■	■		■
DCPS	■		■	■	■		■
NHPS	■			■			■
Nuevo México	■		■		■		■
NYCPS	■			■			
Tennessee	■		■	■	■		

Fuente: Elaboración propia basada en información de Wyckoff y Katz (2018) y CUREE (2018). DCPS = District of Columbia Public Schools; NHPS = New Haven Public Schools; NYCPS = New York City Public Schools.

portafolios docentes, la autoevaluación y los resultados de exámenes estudiantiles (Singapur) y encuestas a estudiantes (República de Corea).

Esta multiplicidad de instrumentos es mucho más común en los sistemas analizados de Estados Unidos, que en sus evaluaciones usan entre tres y cuatro instrumentos (con excepción del NYCPS, que solo incluye dos). De hecho, como se detalla en el gráfico A1 del apéndice, el uso de varios tipos de instrumentos es común a lo largo del país y no se trata de una decisión que únicamente tomaron las autoridades de los seis casos aquí examinados. Salta a la vista que la observación en clase es el instrumento más comúnmente empleado en los sistemas de evaluación docente en Estados Unidos. De hecho, es el único instrumento que está incluido en el 100% de los sistemas (cuadro 3.1 y gráfico A1). También se recurre ampliamente a instrumentos basados en los aprendizajes estudiantiles. En efecto, como se observa en el gráfico A1 del apéndice, el uso de este tipo de información en las evaluaciones docentes es común tanto en los estados como en los grandes distritos del país en general. La variación se presenta en el tipo y la construcción del indicador escogido en cada caso. En el 60% de los distritos considerados por Steinberg y Donaldson (2016) se emplean medidas de valor agregado como indicadores del desempeño docente. Estas se complementan también con medidas de crecimiento en los percentiles estudiantiles y los objetivos de aprendizaje, que —de acuerdo con los autores— son incluidos en el 35% y 39% de los distritos respectivamente. En el caso de los estados, Steinberg y Donaldson encuentran que es mucho más común el uso de medidas de crecimiento en percentiles (59%) y objetivos de aprendizaje (52%) que la utilización de medidas de valor agregado

(30%). Por lo tanto, no es de extrañar que en cuatro de los seis casos aquí analizados se incluyan como instrumentos de la evaluación medidas de desempeño estudiantil en exámenes estandarizados.⁵ Las excepciones tienen lugar en Nuevo México y la ciudad de Nueva York en donde lo que se utiliza como medida de aprendizaje estudiantil es el instrumento asociado a los objetivos de aprendizaje.

Como también se observa en el cuadro 3.1, de los países de alto desempeño el único que incorpora medidas de desempeño estudiantil es Singapur, aunque República de Corea tiene en cuenta a los estudiantes, al incluir encuestas aplicadas a ellos. Este último instrumento se utiliza también en alrededor del 17% de los estados y distritos de Estados Unidos. De hecho, tan solo en los casos de New Haven y de la ciudad de Nueva York no se emplean encuestas estudiantiles dentro de la evaluación docente. Es notorio también que todos los casos analizados para Estados Unidos, excepto el de New Haven, incluyan como instrumento una medida con pocos niveles de estandarización como son los logros de los objetivos de aprendizaje. Finalmente, Denver, DCPS, New Haven y Nuevo México emplean como instrumentos formularios que indagan acerca del nivel de profesionalismo del docente medido en aspectos como su puntualidad, el cumplimiento de sus responsabilidades, el trabajo en equipo, e incluso el cuidado de su apariencia. Es posible que este tipo de instrumento sea muy similar a otros instrumentos utilizados por varios países de la región, los cuales, aunque sin duda aportan alguna información acerca del docente, es poco probable que permitan entender cuáles son sus fortalezas y debilidades disciplinares y pedagógicas, factores que impactan en la calidad de la enseñanza brindada.

Aunque en el caso de Ontario, Reino Unido y Estonia, los sistemas de evaluación incluyen también reuniones con los supervisores, estas, más que instrumentos específicos, se utilizan para obtener información básica del docente, coordinar las actividades de observación y entregar los resultados finales de la evaluación. Por ejemplo, en Ontario (Canadá) el director de la escuela debe realizar una reunión previa a la observación para explicarle al docente los detalles de esta última, y una reunión posterior, donde se analizan los resultados, calificaciones generales y estrategias a seguir. En el caso del Reino Unido, la reglamentación solo exige la observación y una entrevista con el docente en donde se definen los objetivos laborales. Finalmente, como ya se mencionó, para Estonia el sistema de evaluación se centra en el funcionamiento general de la escuela y no en cada docente de manera específica. A pesar de esto, la revisión permitió concluir que acaso la observación sea el instrumento más común utilizado en ese país.

Dada la preponderancia del uso de la observación en clase como instrumento en todos los casos aquí analizados, cabe detallar aspectos relacionados con su diseño e

⁵ Es importante aclarar que la información requerida para la inclusión de medidas de valor agregado o de crecimiento de percentiles estudiantiles en los sistemas de evaluación en Estados Unidos solo está disponible para cerca del 20% de los docentes en el país. Por tanto, aunque se mencione que en un estado o distrito particular se utiliza alguno de estos instrumentos, esto no implica que todos los docentes de dichos sistemas sean evaluados con esa herramienta.

implementación. Como se puntualizó anteriormente, para que esta herramienta pueda brindar información confiable y válida, los detalles de estos dos aspectos son fundamentales. En el diseño, es necesario que las rúbricas de observación sean detalladas y que se basen en claros estándares de desempeño, así como también que se definan con transparencia los actores encargados de llevarlas a cabo, la forma en que se les entregarán los resultados a los docentes y cómo se les dará retroalimentación de sus prácticas en base a ello. Por su parte, en la implementación se requiere un entrenamiento detallado y continuo de los observadores, así como la existencia de protocolos bien definidos con respecto al número y al tipo de observaciones a realizarse (formales/informales; programadas/sorpresa), y en cuanto a las metodologías y los criterios de codificación, los cuales tienen impactos en los puntajes y resultados finales.

Los estudios internacionales dejan en claro que hay una gran variedad en lo que respecta los niveles de rigurosidad relacionados con el diseño y la implementación de los instrumentos de observación en el aula, incluso cuando se comparan distintos casos dentro de Estados Unidos. Al analizar los sistemas de aquellos países que ostentan un alto desempeño, salta a la vista que Estonia es la nación que tiene el menor nivel de centralización y estandarización. Las autoridades educativas no han establecido marcos ni procedimientos obligatorios para llevar a cabo las observaciones. Al tratarse de un sistema tan descentralizado, el proceso y las actividades de la evaluación interna difieren entre una y otra escuela. A este caso le siguen los de Nueva Zelanda, Reino Unido y Ontario, en ese orden, donde, aunque no hay procesos comunes obligatorios para la observación en clase, las autoridades han desarrollado guías para facilitar su implementación. En el caso específico de Nueva Zelanda, la reglamentación exige que se lleven a cabo al menos dos observaciones y al respecto se detallan algunas sugerencias que se deben aplicar.⁶ Finalmente, en los sistemas de Singapur y República de Corea la observación en clase está mucho más detallada y estandarizada. En el primero de estos dos países, por ejemplo, las evaluaciones se llevan a cabo siempre al final del año escolar (octubre) y quedan en manos del supervisor, junto con un panel de superiores que hayan trabajado con el docente (como el director, el subdirector o los jefes de departamento), quienes a través de diferentes protocolos evalúan el potencial del docente con los instrumentos diseñados para tal fin.

Como ya se ha señalado, la varianza en el diseño y la implementación de las observaciones en clase también es importante en Estados Unidos. Donde es evidente que existe una mayor preocupación por el diseño y la estandarización de las observaciones docentes es en el distrito de New Haven. Por ejemplo, el distrito financia y capacita a observadores externos que validan las observaciones docentes que, de acuerdo con su calificación, se concentran en algunas de las colas de la distribución del desempeño. Estos docentes, que por sus calificaciones pueden ser elegibles para oportunidades de liderazgo o bien ser despedidos, tienen por lo tanto la oportunidad de recibir calificaciones libres de sesgos y más

⁶ Véase <https://educationcouncil.org.nz/content/observations>.

exactas. Es interesante también el caso del DCPS, ya que al comienzo exigía que todos los docentes fuesen calificados por observadores externos dos veces al año, pero esta modalidad tan deseable fue eliminada del proceso en 2016 debido a los altos costos monetarios que implicaba.

Al analizar este instrumento se debe también revisar el número de observaciones obligatorias dentro de cada proceso. Como lo demostró el estudio MET (2013), las variaciones en este aspecto afectan de manera importante la confiabilidad de la información recogida. Así, para Estados Unidos se encontró que en un extremo se encuentra Denver, donde se requiere una única observación formal y una observación informal, y en la otra punta se encuentra Tennessee, donde se requieren seis observaciones cada año. En el medio de estos dos casos se hallan la ciudad de Nueva York, que obliga a tener una observación formal y tres informales; los sistemas de DCPS y Nuevo México, donde se requieren un máximo de tres observaciones formales, y New Haven, que demanda dos observaciones formales y tres informales.

Como ya se resaltó, casi la totalidad de los casos analizados en Estados Unidos utilizan como segundo instrumento alguna medida de rendimiento estudiantil. Al igual que sucede con las observaciones en clase, existe una gran variación en el uso de este tipo de información estudiantil dentro de las evaluaciones docentes de Estados Unidos. Aunque DCPS, Tennessee y Nuevo México utilizan medidas de valor agregado de los docentes, su diseño es distinto en cada caso. Por ejemplo, mientras que en DCPS se estima el valor agregado para los docentes de matemáticas e inglés del 4º al 10º grado, en Tennessee y Nuevo México se considera el desempeño estudiantil en una mayor variedad de materias. De manera similar, mientras que en el cálculo del estimador del valor agregado DCPS incluye como controles algunas características socioeconómicas de los estudiantes, Tennessee y Nuevo México no las tienen en cuenta. Las diferencias de estos tres casos en comparación con Denver son aún mayores, ya que el instrumento empleado por este distrito, aunque recurre a información estudiantil, se basa en percentiles de crecimiento en lugar de hacerlo en medidas de valor agregado.

PESOS ASIGNADOS DENTRO DE LA EVALUACIÓN DOCENTE

Para los países que utilizan una multiplicidad de instrumentos en sus observaciones, es importante entender el peso que le asignan a cada uno de ellos, ya que, como lo demostró la investigación de MET (2013), esta decisión tiene impactos importantes en la validez del sistema. La información disponible solo permite llevar a cabo este análisis para el caso de los sistemas de evaluación docente dentro de Estados Unidos. Como se observa en el cuadro A1 del apéndice para el caso general de Estados Unidos y en el cuadro 3.2 para los seis casos particulares aquí analizados, el instrumento de observación en clase es el que en promedio recibe el mayor peso dentro de las evaluaciones docentes de ese país, llegando a más del 50% y el 37% respectivamente. Queda claro así que la observación en clase no solo

CUADRO 3.2 | Pesos asignados a cada instrumento utilizado en las evaluaciones docentes de algunos estados y distritos de Estados Unidos

Distrito/ estado	Observación en clase	Resultados de exámenes a estudiantes	Objetivos de aprendizaje	Encuestas a estudiantes y/o padres	Profesionalismo	Otro
Denver	30%	10%	30%	10%	10%	10%
DCPS	30%	35%	15%	10%	10%	
New Haven	40%		50%		10%	
Nuevo México	25%	50%		5%	15%	5%
NYCPS	50%		50%			
Tennessee	45%	35%	15%	5%		
Promedio	37%	33%	32%	8%	11%	8%

Fuente: Elaboración propia basada en información de Wyckoff y Katz (2018).

Nota: El nivel de profesionalismo del docente se mide en aspectos como su puntualidad, el cumplimiento de sus responsabilidades, el trabajo en equipo e incluso el cuidado de su apariencia.

es el instrumento más comúnmente utilizado sino además el que más peso tiene en las calificaciones finales de las evaluaciones en esa nación. Dado esto, cobra aún más relevancia la discusión anterior en torno a la necesidad de contar con un diseño y una implementación cuidadosos de dicho instrumento, de modo de asegurarse que este logre dar información certera, válida y consistente de las prácticas pedagógicas de los docentes.

Como se detalló en el capítulo anterior, el estudio de MET (2013) sugiere que aquellos instrumentos basados en los resultados de pruebas estandarizadas de conocimientos de los alumnos deben incluirse en las evaluaciones y recibir un tercio del peso en la calificación final cuando sea posible complementarlos con observaciones en clase y encuestas estudiantiles. Esta ponderación permite no solo obtener correlaciones altas con resultados posteriores de los alumnos en distintas pruebas de conocimiento estandarizadas, sino que además es la que otorga la mayor confianza de todas. A pesar de que, como se observó, la mayoría de los estados y distritos de Estados Unidos efectivamente incluyen este instrumento en sus evaluaciones, el peso que se les asigna no es alto. En promedio, Steinberg y Donaldson (2016) encuentran que en dicho país las medidas de valor agregado y de crecimiento en percentiles tienen un peso del 16% y del 10% dentro de los sistemas de evaluación respectivamente (cuadro A1 del apéndice). En los seis casos de Estados Unidos que se detallan en este capítulo, este promedio sube al 37%. Este incremento se explica por el estado de Nuevo México, que a los instrumentos basados en los resultados de pruebas estandarizadas de conocimientos de los alumnos les concede un 50% del peso en sus evaluaciones, y por Tennessee y DCPS, que les otorgan un 35% del puntaje final. Por el contrario, Denver tan solo asigna el 10% del puntaje final a estas medidas de desempeño estudiantil.

De hecho, al comparar el peso asignado a los distintos instrumentos en los seis casos examinados para Estados Unidos, resalta el hecho de que el instrumento con el segundo mayor peso es aquel relacionado con el análisis del cumplimiento de las metas

de aprendizaje que ha fijado el docente al comienzo del año. Por ejemplo, New Haven y la ciudad de Nueva York le asignan el 50% de la calificación final a este instrumento y, de los seis casos considerados, tan solo Nuevo México no lo incluye dentro de la evaluación docente. La preponderancia concedida a este instrumento iría en contra de las recomendaciones de expertos en el tema, ya que, por su diseño e implementación, es una medida mucho más subjetiva y difícilmente estandarizable. Sin embargo, cabe aclarar que estos seis casos parecerían atípicos en relación con lo que ocurre en Estados Unidos, ya que, como se observa en el cuadro A1 del apéndice, el promedio a nivel nacional en cuanto al peso que se le asigna a este instrumento baja al 10%.

El último aspecto que deben decidir las autoridades educativas en torno al puntaje dentro de la evaluación docente es el tipo de calificación que se le asignará a cada docente y las consecuencias que de allí se derivan. Con respecto al primer aspecto, de acuerdo con el estudio de Kraft y Gilmour (2017), la mayoría de los estados optó por tener una calificación con cuatro categorías: Muy Efectivo, Efectivo, En Desarrollo y No efectivo. De hecho, únicamente tres de los 24 estados considerados en el mencionado estudio incluyeron una quinta categoría de “Excepcional”. Sin embargo, a pesar de los avances en la utilización de múltiples instrumentos para la evaluación docente, acordes con la labor compleja y las múltiples dimensiones que buscan analizar, y los esfuerzos en tiempo y dinero que esto ha traído consigo, la calificación final otorgada a los docentes en Estados Unidos sigue concentrada en la cola superior de la distribución. En la mayoría de los estados analizados menos del 10% de los docentes fue clasificado en las últimas dos categorías (véase el gráfico A2 del apéndice).⁷ Tal como lo mencionan Wyckoff y Katz (2018), esta concentración en la cola superior no es consistente con otras medidas de la calidad educativa, como —por ejemplo— los resultados que obtienen los estudiantes en pruebas estandarizadas. De acuerdo con los autores, sería de esperar que las calificaciones de los docentes siguieran en parte la distribución de los alumnos. Las diferencias entre ambos llevan a concluir que, aunque sin duda hay una mejoría en los sistemas de evaluación actuales al compararlos con los que existían antes, son necesarios mayores esfuerzos para continuar mejorándolos.

CONSECUENCIAS LIGADAS A LOS RESULTADOS DENTRO DE LA EVALUACIÓN

Como quedó claro en la teoría del cambio (gráfico 2.1), el uso que se les dé a los resultados obtenidos en las evaluaciones docentes es fundamental para lograr mejoras efectivas

7 La revisión de CUREE (2018) no detalla estos aspectos y solo menciona como es la clasificación final otorgada para tres países en su análisis. En Ontario únicamente existen dos categorías (satisfactorio y no satisfactorio); en Singapur el sistema otorga un puntaje numérico de 1 a 100 y en base a este los docentes son clasificados con un desempeño de excelente a pobre; en República de Corea los docentes son catalogados en una de cinco categorías que van de la A (la mayor) a la E (la menor). No hay información con respecto a la distribución de la calificación final obtenida por los docentes para ninguno de los sistemas educativos analizados por los autores.

CUADRO 3.3 | Consecuencias ligadas a los resultados de la evaluación docente en los sistemas educativos exitosos seleccionados

País	Diagnóstica	Bonificaciones/ salarios	Ascenso	Formación	Permanencia
Ontario (Canadá)					
Reino Unido					
Nueva Zelanda					
Singapur					
Rep. de Corea					
Estonia					
Denver					
DCPS					
New Haven					
Nuevo México					
NYCPS					
Tennessee					
Uso promedio	17%	75%	67%	25%	25%

Fuente: Elaboración propia basada en información de Wyckoff y Katz (2018) y CUREE (2018).

en el cuerpo docente y en la calidad educativa recibida por los estudiantes. En el cuadro 3.3 se exhiben las consecuencias de la calificación final obtenida por los docentes en los sistemas de evaluación en términos de su carrera profesional, los salarios o bonificaciones que devengan y las posibilidades de formación continua. Como ya se detalló en el capítulo anterior, aunque existe evidencia de que la sola implementación de la evaluación puede tener impactos positivos (véase el trabajo de Taylor y Tyler, 2012), lo fundamental es entender cómo se utilizan los resultados que se desprenden de los sistemas de evaluación. Las consecuencias más comunes atadas a los resultados de las evaluaciones en los 12 casos analizados tienen que ver con los incrementos salariales o las bonificaciones y las posibilidades de ascenso. Solo hay tres casos en los cuales los resultados de las evaluaciones se vinculan bien sea a aspectos del desarrollo profesional o bien a la estabilidad laboral. Esto implica, respectivamente, que los docentes que obtengan bajos resultados en sus evaluaciones quedan sujetos a programas de formación o incluso a la remoción del sistema cuando estos resultados se obtienen de manera consecutiva.

De los sistemas de alto desempeño educativo tan solo en el caso de Ontario la evaluación cumple como única función un rol exclusivamente diagnóstico, que permite identificar el alcance de los objetivos de desarrollo profesional de los docentes. En Reino Unido y Estonia, por ejemplo, los resultados de la evaluación son el insumo principal con el que se discuten las necesidades de desarrollo profesional, la progresión de la carrera y por supuesto los niveles salariales que se pueden devengar. De manera similar, en Nueva Zelanda la clasificación final está vinculada a la progresión salarial y a la revisión anual de los objetivos profesionales. En

República de Corea, la evaluación tiene consecuencias tanto aditivas como formativas, en la medida en que los resultados buscan identificar docentes de alto y bajo rendimiento para brindarles acceso a oportunidades de investigación o a programas de desarrollo profesional, respectivamente. Por último, los resultados de los tres sistemas de evaluación de Singapur definen aspectos relacionados con promociones dentro de la carrera, bonificaciones de desempeño anual e incluso con la posibilidad de despido para aquellos que no cumplan con los estándares requeridos en la profesión. Vale la pena resaltar la importancia que se les otorga en este país a los sistemas de formación en servicio, lo cual se ve reflejado en el hecho de que el Ministerio de Educación financia hasta 100 horas de desarrollo profesional. Si bien estas horas de capacitación no están totalmente atadas a los resultados de la evaluación, los resultados de la misma se hallan directamente ligados a estas actividades, ya que los jefes de departamento de cada escuela los utilizan para definir las necesidades de cada docente y la organización de talleres dentro de las instituciones.

En lo que respecta a los sistemas analizados para Estados Unidos, el único caso que no ata los resultados de las evaluaciones a aspectos salariales directamente es el distrito de la ciudad de Nueva York. Denver, DCPS y New Haven vinculan los resultados a bonificaciones y salarios tanto de forma positiva (incrementos) como negativa (congelamiento); Nuevo México lo hace tan solo a aspectos positivos (bonificaciones) y Tennessee tan solo a aspectos negativos (congelamiento). Los resultados obtenidos por los docentes en Denver, DCPS, New Haven y NYCPS definen sus posibilidades de ascenso, y para DCPS y New Haven incluso su permanencia dentro del magisterio.

El único caso en el cual los resultados se asocian a planes de mejoramiento y formación es el del distrito de New Haven. De acuerdo con lo establecido por ley, en Nuevo México y Tennessee, al igual que en la ciudad de Nueva York, los resultados de las evaluaciones docentes deberían informar acerca de los planes de desarrollo profesional, pero no hay evidencia de que en dichos sistemas existan efectivamente planes holísticos para el desarrollo profesional de los docentes. Por el contrario, en New Haven sí se halla evidencia de que todos los docentes llevan a cabo un plan de desarrollo con metas específicas. En particular, a lo largo del año, los docentes pueden entrar en un “Plan de Apoyo Estructurado”, y aquellos que hayan obtenido rendimientos bajos en sus evaluaciones ingresan en un programa más integral llamado “Plan de Mejoramiento Intensivo”.⁸ Al comparar estas consecuencias con las encontradas para el mucho más amplio grupo de estados y distritos

⁸ Como lo detallan Wyckoff y Katz (2018), algunos de los sistemas analizados cuentan con planes o programas que buscan apoyar el desarrollo profesional de los docentes. Sin embargo, estos no están necesariamente ligados a los sistemas de evaluaciones docentes. Por ejemplo, Tennessee llevó a cabo un programa piloto de un modelo de colaboración entre pares en la escuela; a partir de datos de observación en el aula, los docentes de bajo rendimiento se asociaron con docentes de alto rendimiento con el objetivo de que trabajaran juntos durante el año escolar 2013-14. Una evaluación de impacto del mencionado piloto sugiere que la estrategia aumentó los logros estudiantiles (Papay et al., 2016) y es posible que el estado decida ampliar el programa. DCPS también cuenta con un programa conocido como

analizados por Steinberg y Donaldson (2016), y que se resumen en el cuadro A2 del apéndice, saltan a la vista diferencias notables. De hecho, el desarrollo profesional es el aspecto más comúnmente ligado a los resultados de las evaluaciones de desempeño según estos investigadores. Le sigue como consecuencia más común la terminación de los contratos para aquellos docentes que muestren un bajo desempeño, y finalmente, en menos de la mitad de los sistemas analizados los resultados se atan a ascensos o bonificaciones (la única excepción es para los “estados de origen”, home states, en donde en el 75% de los casos analizados sí se vinculan los resultados de las evaluaciones docentes con aspectos de progresión en la carrera).

Del análisis de las principales características de los sistemas de evaluación docente en los 12 sistemas educativos aquí considerados, se desprenden varias conclusiones clave.

La primera es que, aunque no existe un único diseño ideal, la gran mayoría de los sistemas se ha movido hacia el uso de evaluaciones que cuentan con una multiplicidad de instrumentos orientados a entender las dimensiones del desempeño docente que los países consideran importantes.

En segundo lugar, aunque todos incluyen la observación en clase como herramienta central en los procesos de evaluación, su diseño e implementación varía de manera considerable. Como lo han demostrado las investigaciones alrededor del tema, esto genera impactos en la validez y confiabilidad de los instrumentos. Por lo tanto, asegurar la implementación llevada a cabo en sistemas como el de New Haven, donde se utilizan observadores externos entrenados que realizan al menos dos observaciones por docente, son ejemplos deseables, pero sin duda costosos.

En tercer lugar, el uso de instrumentos basados en resultados obtenidos por los propios estudiantes en exámenes estandarizados parece ser una herramienta mucho más común en los sistemas educativos de Estados Unidos que en el resto de los países estudiados. La explicación de este fenómeno puede explicarse en parte por el financiamiento que el gobierno federal de ese país ha dado a sus sistemas educativos recientemente para la implementación de dichos exámenes en múltiples grados y materias. La evidencia sugiere que este tipo de herramientas quizá sea el que mayor validez posee para identificar la efectividad de los docentes en el conocimiento disciplinar adquirido por los estudiantes, medido por estas pruebas específicas, pero queda claro que requiere inversiones iniciales importantes que hoy en día no existen en los países de América Latina y el Caribe y que dejarían a la mayor parte de los docentes sin información en este aspecto.

Por último, el uso de los resultados que surgen de las evaluaciones es fundamental. Curiosamente, los casos analizados en este capítulo sugieren que la mayoría de los sistemas

LEAP, dirigido a reforzar las habilidades de los docentes a través de grupos pequeños donde se discuten aspectos relacionados con la planificación, las estrategias de enseñanza, los avances de los estudiantes e incluso temas disciplinares. Sin embargo, ninguno de estos programas, por ahora, está atado directamente a los sistemas de evaluación docente.

ata a un mismo tipo de evaluación tanto las consecuencias formativas como aquellas relacionadas con incentivos monetarios y de carrera. Esto contrasta de manera notable con algunos casos de América Latina donde los sistemas han evolucionado hacia la implementación de distintos tipos de evaluaciones para alcanzar diferentes propósitos, que es lo que sucede en República de Corea.



CAPÍTULO 4

SITUACIÓN DE LOS SISTEMAS DE EVALUACIÓN DOCENTE EN AMÉRICA LATINA Y EL CARIBE

Varios países de América Latina y el Caribe están llevando a cabo importantes esfuerzos para desarrollar sus sistemas de evaluación docente. Este capítulo analiza y sistematiza la información recolectada sobre estos sistemas, lo cual permite una mirada holística al estado de los sistemas de evaluación en la región y un análisis de las similitudes y las diferencias en el tipo de evaluaciones docentes que se están implementando. La información de cada país se reunió a partir de una revisión de los documentos oficiales y de encuentros con los encargados de evaluación docente en los respectivos ministerios.¹ Este capítulo, junto con la evidencia presentada en los capítulos anteriores, permitirá entender y contextualizar los avances alcanzados por los países de la región y guiar aspectos relevantes de las políticas asociadas a este eje fundamental de la carrera docente en el futuro.

Los 19 sistemas educativos analizados que forman parte de la presente sistematización son los de provincia de Buenos Aires (Argentina), Belice, Bolivia, Santa Catarina (Brasil), Chile, Colombia, Costa Rica, Ecuador, El Salvador, Guatemala, Honduras, Jamaica, México, Panamá, Paraguay, Perú, República Dominicana, Trinidad y Tobago y Uruguay. Para todos ellos, la información corresponde a las prácticas de evaluación docente vigentes en los años más recientes.²

El capítulo analiza cuatro tipos de evaluaciones docentes (véase el gráfico 4.1) que se realizan en tres momentos distintos de la profesión: i) evaluaciones para el ingreso a la carrera, ii) evaluaciones durante el período de prueba o al finalizar el mismo y

¹ La información presentada se nutre también del documento de antecedentes producido por Rodríguez y Rivera (2018).

² La información se recopiló durante 2017, y presenta actualizaciones a 2019 en la medida de lo posible.

GRÁFICO 4.1 | Tipos de evaluaciones docentes analizadas en el estudio

Fuente: Elaboración propia sobre la base de Rodríguez y Rivera (2018).

iii) evaluaciones para docentes en ejercicio. Mientras en las primeras dos etapas solo existe un tipo de evaluación específica, en la tercera etapa se consideran dos evaluaciones diferentes con características particulares: evaluaciones obligatorias periódicas y evaluaciones para ascensos o bonificaciones. Para cada tipo de evaluación, se revisa si existe mención en los estatutos, investiga si efectivamente se implementa, compara los instrumentos utilizados y cómo los resultados obtenidos están atados o no a procesos de formación continua y si determinan implicaciones a lo largo de la carrera docente.

Debido a que las evaluaciones se aplican en distintas etapas de la carrera, sus objetivos, su diseño y su implementación tienen aspectos particulares y varían incluso dentro de un mismo país. La primera evaluación analizada, que se realiza en la gran mayoría de los sistemas educativos de la región que integran este estudio excepto en Belice y Jamaica, define si un individuo puede entrar o no en la fuerza docente del sistema público. En algunos países este tipo de evaluación se conoce como "concurso docente". La segunda evaluación, que es la menos común de todas en los sistemas analizados, determina si docentes principiantes que ingresaron en el magisterio tienen las habilidades pedagógicas y disciplinares requeridas para asegurar el aprendizaje de sus estudiantes y, por lo tanto, para permanecer en el sistema y en la carrera docente. Finalmente, en la última etapa se analizan dos tipos de evaluaciones que se administran a docentes que superaron el período de prueba con éxito, han sido nombrados en un cargo docente y por tanto se encuentran regidos por los estatutos docentes de cada país. Las primeras (a las que en este estudio se denomina *evaluaciones obligatorias periódicas*) se aplican a docentes de manera obligatoria en períodos establecidos. De acuerdo con los estatutos docentes, en la mayoría de los casos los resultados de estas evaluaciones determinan la permanencia o estabilidad laboral y, en ocasiones, también las capacitaciones que los docentes deberían recibir para subsanar falencias. Las segundas

(a las que en este estudio se denomina *evaluaciones para ascensos horizontales*) se implementan para establecer si el docente puede avanzar o no en su carrera profesional.

La sistematización permite extraer conclusiones importantes de cada uno de los cuatro tipos de evaluaciones analizados en el presente estudio. Respecto de las evaluaciones de ingreso a la carrera, la legislación de 17 de los sistemas educativos incluidos en este trabajo las menciona de manera explícita y las hace obligatorias para la entrada al magisterio. No obstante, se encontró que en cinco de estos sistemas las evaluaciones no se realizan. En estos casos, podrían ocurrir tres situaciones: i) la selección se basa únicamente en la verificación de títulos académicos, ii) el proceso puede ser corrupto, iii) no existe información que compruebe de forma efectiva si las evaluaciones se están llevando a cabo y en qué condiciones (FEREMA, 2015). En la gran mayoría de los 12 sistemas restantes, la selección se basa en la evaluación del conocimiento disciplinar y pedagógico de los aspirantes a través de la presentación de exámenes como principal instrumento. En casi la mitad de ellos, los exámenes se complementan con entrevistas. Sin embargo, no es posible conocer su estructura y su nivel de profundidad.

Las falencias en la calidad de la formación inicial y las características de los interesados en ingresar a la profesión docente en la región (reseñadas por Elacqua et al., 2018), sumadas a la dificultad que tienen los ministerios de educación para remover del cargo a docentes con bajos niveles de desempeño, convierten a las evaluaciones de ingreso a la carrera en un elemento esencial para identificar mejor las características y capacidades de los futuros docentes. En ese sentido, continuar con los procesos de mejora y de implementación adecuada de esta evaluación podría traer beneficios en la calidad de la fuerza docente en la región. Aunque la evidencia aún es escasa, estudios para Colombia (Ome, 2012; Brutti y Sánchez, 2017) y México (Estrada, 2019) sugieren que estos sistemas de selección (con evaluaciones para ingresar a la carrera) están teniendo algunos impactos positivos en la calidad educativa de los países que los implementan. Asimismo, las investigaciones disponibles indican que el nivel de motivación también es fundamental para un desempeño adecuado en el sector público (Banuri, Keefer y De Walque, 2018), y la docencia no debe ser la excepción. Por eso, los procesos de selección de docentes para la carrera magisterial deberían tener también en consideración este aspecto para atraer a candidatos motivados (Elacqua et al., 2018).

De manera similar, los primeros años de ejercicio en la carrera docente tienen un papel sustancial en el desarrollo de las capacidades de los docentes (Jensen et al., 2012; Ávalos, 2016). Actualmente, la mayoría de los sistemas analizados contempla en la legislación algún tipo de programa de inducción docente y lo implementa. Aunque la calidad de los programas no resulta clara, su existencia es, en sí misma, un paso importante. Hay evidencia de que la correcta inserción de los docentes noveles y el apoyo a los recién ingresados en la carrera están en la discusión política en un poco más de la mitad de los países de la región.³ Aun así,

³ El diseño, la implementación y los impactos de estos programas se encuentran fuera del alcance de este estudio. Los lectores pueden acceder a información destacada en documentos como Ávalos (2016) y Marcelo y Vaillant (2017), entre otros.

la revisión de los sistemas de evaluación de docentes en esta etapa de la carrera demuestra que la evaluación al finalizar el período de prueba es la menos común de todas.

En 13 sistemas educativos de los 19 sistemas analizados en este trabajo no se lleva a cabo ninguna evaluación en esta etapa de la profesión, a pesar de que la evaluación está explícitamente mencionada en la legislación en varios de ellos. Si bien en tres sistemas se realiza una evaluación, esta se basa en un único instrumento, que consiste en un formulario completado por el director y el docente. Debido a su estructura, esta evaluación no permite detectar aspectos de la práctica docente que podrían brindar información útil para asegurar una correcta inserción de los docentes en el sistema. Solo en tres sistemas (Santa Catarina en Brasil, Chile y México) la evaluación de docentes principiantes se basa en instrumentos múltiples que acreditan realmente su desempeño. En el caso de Santa Catarina (Brasil) se trata de pruebas diseñadas de manera específica para tal fin, mientras que en Chile y México los docentes que cumplen dos años de ejercicio deben efectuar las mismas pruebas obligatorias periódicas que hacen los demás miembros de la fuerza docente (las características de estas pruebas se describen en detalle más adelante). Por otra parte, a pesar de que en los sistemas donde se aplican estas pruebas los resultados están atados a la posibilidad de continuar o no ejerciendo en la fuerza docente del país, la realidad es que pocos docentes son retirados del magisterio por bajo desempeño incluso en estas etapas tempranas de la carrera.

En 17 sistemas educativos de los 19 sistemas analizados la legislación establece que luego de ser nombrados en su cargo y de pasar el período de prueba los docentes deben realizar una evaluación obligatoria periódica. En 11 de estos sistemas la periodicidad de esta evaluación es anual mientras que en los seis sistemas restantes la periodicidad varía entre tres y cuatro años dependiendo del país y de los resultados que los docentes obtengan en la evaluación. Esta división, aunque parece sutil, conlleva inmensas diferencias en el tipo de evaluación reglamentada en la legislación e implementada en el sistema. La gran mayoría de los estatutos que obligan a los docentes a realizar una evaluación anual definen que estas se basan en un único instrumento compuesto por formularios preestablecidos que el director o el superior jerárquico deben diligenciar.⁴ Solo en cinco de los sistemas analizados de este grupo se encontró evidencia de que esta evaluación efectivamente ocurre. Más aún, al basarse en un instrumento sencillo que no permite recoger evidencia del desempeño del docente, la mayoría de los docentes se cataloga con desempeños sobresalientes y son pocas o nulas las consecuencias que los resultados conllevan en el desarrollo profesional de los profesores. A su vez, esta cantidad notable de docentes catalogados con desempeños sobresalientes o buenos implica que la desvinculación de docentes poco

⁴ Como se detalla en la sección respectiva, la única excepción a esta tendencia se da en Paraguay, país que en 2017 introdujo una evaluación con múltiples instrumentos. Como el sistema aún se encuentra en la fase inicial y se espera que su implementación sea gradual, es posible que la periodicidad deba modificarse en un futuro para que su aplicación sea viable.

efectivos del magisterio es casi inexistente. Quizás, el mayor uso que se les da a los resultados que surgen de las evaluaciones tiene que ver con los ascensos en el escalafón, que debido a la estructura de las evaluaciones terminan dependiendo más de los años de experiencia y de los títulos que de las destrezas y habilidades docentes.

En los sistemas educativos que reglamentan evaluaciones obligatorias con una periodicidad mayor la situación es distinta. En primer lugar, en la totalidad de estos sistemas las evaluaciones efectivamente se implementan, aunque no en todos los casos se efectúan tal como lo establece la legislación. En segundo término, por lo general los sistemas incluyen múltiples instrumentos en busca de evaluar de manera efectiva el desempeño de sus docentes. Aunque la elección final de instrumentos es variada, la observación en clase y la presentación de exámenes estandarizados son los más utilizados y los que mayor peso han recibido dentro de la evaluación por parte de las autoridades educativas. En tercer lugar, la gran mayoría de los sistemas contempla el uso de evaluadores externos a la institución educativa para la implementación de las evaluaciones. En varios de ellos, incluso, existen instituciones dedicadas a esta labor, lo cual le otorga al sistema una independencia mayor que la que tiene cuando la evaluación la realiza solo el director. Por último, aunque en la mayoría de estos sistemas la legislación ata los resultados de las evaluaciones a decisiones de permanencia y/o de programas de formación en servicio, los sistemas no han logrado aún implementar esto de manera profunda. Por lo que se ha observado en el marco de esta revisión, se producen pocos despidos efectivos de docentes del magisterio como consecuencia de esto, al igual que en los sistemas con evaluación obligatoria anual. Por otro lado, a pesar de que Chile, México y Perú atan los resultados a este tipo de programas solo se están aplicando de manera efectiva en el primer país.

El último tipo de evaluación analizada es la que los sistemas implementan para permitir los ascensos en la carrera docente. En particular, el capítulo estudia las posibilidades y los requisitos que exige la legislación para ascender en la carrera y acceder a mayores niveles de ingreso, prestigio y, quizás, responsabilidades, sin la necesidad de dejar las aulas para ocupar cargos más administrativos. Estos ascensos, que como se explicó anteriormente se denominan ascensos horizontales, son distintos de aquellos en los que los docentes ascienden a cargos administrativos como directores o coordinadores, que se llaman ascensos verticales. La revisión permitió dividir los sistemas educativos en tres grupos claros de acuerdo con las posibilidades y los requisitos de ascensos horizontales: i) sistemas donde no existe este tipo de ascensos, ii) sistemas donde existe la posibilidad de ascender horizontalmente pero estos ascensos dependen de forma exclusiva de los años de experiencia y/o los títulos académicos y iii) sistemas en los cuales los ascensos horizontales son meritocráticos y es preciso superar pruebas específicas para alcanzarlos. En cuanto a este último grupo es importante resaltar tres puntos. En primer lugar, el tipo de evaluación al que están sujetos los ascensos horizontales y las características asociadas, tales como la periodicidad, los instrumentos y los responsables, varían de manera significativa entre sistemas. En segundo término, la revisión encontró que en algunos sistemas los docentes no deben presentar

una evaluación adicional a la obligatoria, sino que los puntajes obtenidos en la evaluación obligatoria son tenidos en cuenta en las decisiones de ascensos horizontales. En general, los sistemas que llevan a cabo evaluaciones obligatorias con una multiplicidad de instrumentos aprovechan la información brindada por estas pruebas como insumo esencial al momento de tomar decisiones de ascenso. En otros sistemas, aunque la reglamentación es clara respecto del requisito de una evaluación específica de ascensos horizontales, no existe evidencia que permita verificar que efectivamente lo escrito en la ley está siendo aplicado. Por último, como se ha observado en otras evaluaciones, existen diferencias importantes entre sistemas con relación al nivel de implementación de los requisitos e, incluso, dentro del mismo sistema la situación *de facto* se opone en ocasiones a la *de jure*.

A continuación se presenta el análisis de dos cuestiones que están muy ligadas a la evaluación docente y a la carrera en general: i) la existencia de estándares de desempeño docente que guíen no solo la estructura de las evaluaciones docentes implementadas sino también temas trascendentales como la formación inicial, la formación en servicio y las propias expectativas de la fuerza docente en el país; ii) la reglamentación de la carrera docente, sobre todo en lo relacionado con la evaluación docente y la manera en que esta determina, o no, las posibilidades de permanencia, capacitación, bonificaciones y ascensos en cada sistema educativo. En varios de los sistemas educativos analizados, las evaluaciones docentes se aplican en el marco de leyes de carrera docente recientes, que buscan hacer la carrera más meritocrática y elevar el prestigio de la profesión. Las nuevas carreras les permiten a los docentes acceder a mayores salarios y a mejores oportunidades de crecimiento profesional. Luego, el capítulo continúa con toda la información recolectada y estudiada sobre evaluación docente y la descripción detallada de aspectos esenciales, también relevados para la revisión de las experiencias por fuera de la región, como los instrumentos utilizados, los pesos asignados, las consecuencias que emanan de los resultados y los actores involucrados en las distintas etapas, entre otras características que la literatura especializada en el tema considera fundamentales.

Finalmente, aunque no es un tema que se analice en detalle en este libro, es importante resaltar que, en la mayoría de los países de la región, la implementación inicial de las nuevas evaluaciones fue rechazada por los sindicatos de maestros, por lo que numerosos gobiernos debieron sortear (y en algunos casos siguen sorteando) paros nacionales importantes de la fuerza docente en oposición a la aplicación de evaluaciones. Para reducir dicha resistencia, los gobiernos han utilizado una variedad de estrategias, como, por ejemplo, atar los resultados únicamente a consecuencias positivas para los docentes, publicar los bajos resultados de aprendizaje de los alumnos para dar un mayor sentido de urgencia a la implementación de reformas, movilizar el apoyo de la opinión pública a los procesos de evaluación, y, en algunos casos, establecer penalidades para los docentes que no participan en las evaluaciones, entre otras. El manejo político que hace falta para llevar a cabo estas evaluaciones al igual que el gran número de docentes del sector público que hay en muchos países de la región han provocado que la implementación se esté realizando de

forma gradual (Elacqua et al., 2018). Algunos estudios que han investigado aspectos vinculados con los acuerdos y la implementación final de programas y políticas docentes en la región incluyen los de Mizala y Schneider (2014 y 2019), en los que los autores analizan la economía política alrededor de las reformas realizadas en el sistema de evaluación docente y la reglamentación de la carrera docente en Chile, respectivamente. Por su parte, Bruns, De Gregorio y Taut (2016) estudian el tema de manera general con un énfasis en particular para la región latinoamericana.

A pesar de las dificultades y aunque todavía es necesario realizar más acciones en este eje de la profesión, los avances alcanzados en varios países de la región en cuanto a la evaluación docente han sido positivos, tal como se describe a continuación.

LOS ESTÁNDARES DE DESEMPEÑO DOCENTE EN LA REGIÓN

La existencia y la implementación de estándares de desempeño docente en cada país constituyen pasos previos necesarios para entender el alcance y la efectividad de los sistemas de evaluación docente. Los estándares son una herramienta transversal a los tres tipos de evaluaciones docentes y un instrumento esencial para guiar y fundamentar todos los ejes dentro de la profesión docente. Los estándares son diferentes en cada país, pero de manera general incluyen dimensiones para asegurar que los docentes cuenten con conocimientos disciplinares, curriculares. Dentro de cada país, lo que se espera de los docentes puede variar según el contexto y la realidad de los alumnos, por eso cada sistema debe definir indicadores específicos por nivel, modalidad, idioma de enseñanza, entre otros.

Como lo enfatiza la literatura mencionada anteriormente, un sistema de evaluación docente debe ser válido, confiable y transparente para que los docentes lo acepten y para que los esfuerzos que se requieren para su diseño y su implementación puedan generar los resultados esperados. Para alcanzar este objetivo es necesario que todos los actores del sistema educativo tengan claro qué es lo que un docente debe *saber* y debe *saber hacer*. La formación inicial y la formación en servicio deben ser diseñadas para asegurar que los futuros docentes y los que están en ejercicio logren efectivamente estos estándares en su desempeño diario. De igual manera, se debe seguir una lógica en el diseño y la implementación de los instrumentos de evaluación que permita identificar si un individuo que ejerce la profesión cumple los estándares de desempeño de manera adecuada. Por último, los docentes mismos deben conocer lo que el sistema espera de ellos y las características que les permitirán ser mejores profesionales. Todo esto se logra solo cuando se tienen estándares de desempeño definidos, difundidos y utilizados por todos.

A pesar de su relevancia, pocos países en la región han diseñado esta herramienta y la han incorporado al sistema educativo, en contraste con lo que ocurre en países desarrollados o en sistemas educativos exitosos. El gráfico 4.2 muestra que solo en cinco países existen y se utilizan estándares de desempeño, con distinta intensidad: Chile, Ecuador, México, Perú y República Dominicana. En estos cinco países los estándares de desempeño

GRÁFICO 4.2 | Avance de los estándares de desempeño docente en América Latina y el Caribe



Fuente: Elaboración propia sobre la base de Rodríguez y Rivera (2018).

han sido discutidos y aprobados por diversos sectores de la sociedad, como ministerios, docentes y en algunos casos organizaciones no gubernamentales del sector educativo, y guían el diseño y la aplicación de la política pública docente. En la actualidad, Guatemala y Panamá están en proceso de elaboración de los estándares; mientras que Colombia, Honduras, Jamaica y Paraguay incorporaron ciertos estándares específicos para sus procesos de evaluación docente desarrollados recientemente. Los demás sistemas educativos no cuentan con estándares de desempeño docente. Algunos, como Bolivia y El Salvador, han definido perfiles de egreso de las carreras de formación docente y otros tienen estatutos docentes que describen de manera general las obligaciones mínimas que deben cumplir los docentes que se encuentran regidos por ellos.

Dentro del grupo de países que han llevado a cabo el proceso de discusión, diseño, aprobación e implementación de estándares docentes en la región, Chile es el pionero. Este país creó en 2003 el Marco para la Buena Enseñanza, basado en cuatro dominios: la preparación para la enseñanza, la creación de un ambiente propicio en clase, la enseñanza para el aprendizaje de todos los estudiantes y las responsabilidades profesionales. Desde su creación, el Marco ha orientado las políticas de fortalecimiento de la profesión docente, ha sentado las bases para el diseño de programas de formación inicial y en ejercicio y ha sido utilizado en los procesos de evaluación del desempeño. Más importante aún es que ha permitido que los docentes del país tengan una guía para reflexionar y evaluar sus propias prácticas de enseñanza y educación. De hecho, entre 2016 y 2017 se realizó una consulta para actualizarlo sobre la base de nuevas discusiones y procesos participativos con la academia y los propios docentes. La propuesta fue presentada al Consejo Nacional de Educación, ente que deberá revisarlo y aprobarlo antes de que entre en vigor la nueva versión.

En los demás países que integran este grupo los estándares de desempeño son más recientes y, por tanto, su utilización ha sido menor. Sin embargo, en ellos también se llevaron a cabo procesos reflexivos que permitieron establecer con claridad lo que se espera de la labor docente. En el caso de Ecuador, los estándares de 2012 constan de cuatro dimensiones referentes al dominio disciplinar y curricular, la gestión del aprendizaje, el desarrollo profesional y el compromiso ético. Además, cada una de estas cuatro dimensiones tiene estándares generales, estándares específicos e indicadores. Por su parte, en México los estándares definen que un buen docente debe: i) conocer a sus alumnos; ii) organizar, evaluar su trabajo y hacer una intervención didáctica; iii) mejorar continuamente su formación y sus prácticas pedagógicas; iv) asumir sus responsabilidades legales y éticas, y v) participar de manera activa en la escuela y en la comunidad. En Perú, el Marco de Buen Desempeño Docente aprobado en 2012 contempla cuatro dominios que incluyen la preparación para el aprendizaje de los estudiantes, la enseñanza para el aprendizaje de los estudiantes, la participación en la gestión de la escuela articulada a la comunidad y el desarrollo de la profesionalidad y la identidad docente. Finalmente, los estándares de República Dominicana aprobados en 2014 también comprenden cuatro dimensiones: i) el reconocimiento del estudiante y sus individualidades como centro de atención, ii) el conocimiento

del contenido disciplinar y curricular, iii) la implementación de procesos de enseñanza y aprendizaje efectivos y iv) el compromiso personal y profesional.

Las diferencias entre los estándares de desempeño en la región, evidentes incluso en el número de dimensiones contempladas en cada uno, confirman la complejidad de la labor docente. Asimismo, dichas diferencias demuestran que ni siquiera es posible establecer en un marco común lo que se espera de un docente de excelencia, a fin de que pudiera ser aplicado de forma general en aquellos países de la región que aún no lo tienen. Es necesario que los sistemas educativos que no cuentan con esta herramienta definan lo que esperan de su cuerpo docente, de manera consensuada y de acuerdo con su contexto y sus objetivos de educación. Además de fijar los estándares de desempeño es esencial que una vez diseñados logren guiar los distintos ejes de la carrera docente, incluido el de la evaluación docente. Para esto es crucial que en las discusiones de diseño participen activamente los docentes y se construyan consensos acerca de lo que es relevante en la práctica docente en cada sistema, tal como sucedió con los cinco países que ya tienen los estándares definidos. Solo a través de estos acuerdos se puede generar el ambiente propicio que facilite su efectiva incorporación a cada etapa de la carrera docente y del sistema educativo en general. Por otra parte, es importante mencionar que la vaguedad en algunos de estos estándares representa un reto a la hora de medirlos de manera efectiva a través de indicadores específicos y de una evaluación docente.

LA REGLAMENTACIÓN DE LA PROFESIÓN EN LA REGIÓN

El segundo aspecto íntimamente relacionado con la estructura de las evaluaciones docentes en cada país lo constituyen los estatutos docentes que rigen la carrera y reglamentan dichas evaluaciones. Como lo detalla el estudio de la Unesco (2015), la reglamentación de las carreras docentes en la región es muy diversa en aspectos que incluyen, entre otros, el nivel legislativo en el que está consignada, el momento en que fue redactada, pero sobre todo en el tipo de carrera que de ella se desprende. En cuanto al nivel legislativo, en algunos países la reglamentación de la carrera está ligada a la Constitución, en otros tiene forma de estatutos específicos únicos que rigen la profesión, mientras que en un tercer grupo la reglamentación está incluida dentro de las leyes generales de educación. De igual forma, en países federales como Argentina y Brasil existen estatutos específicos para cada ente administrativo; en tanto que en los países centralizados hay un único marco legal que todos deben cumplir.

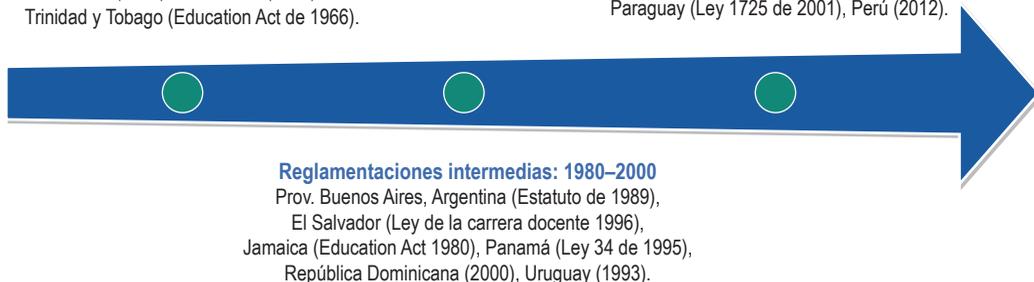
El espacio de tiempo en el que la reglamentación de la carrera fue establecida en los países de la región es amplio y explica en parte las diferencias en la visión de la profesión que expresa la reglamentación en cada país. Por ejemplo, como se observa en el gráfico 4.3, en la región coexisten estatutos docentes desarrollados en los años cincuenta, como el de Bolivia, de 1957, con estatutos muchos más recientes como el de Chile, reformulado en 2016. En el gráfico se divide a los sistemas educativos incluidos en esta revisión

GRÁFICO 4.3 | Reglamentación de la carrera docente en la región a lo largo del tiempo**Primeras reglamentaciones en la región que aun siguen vigentes: 1957–1979**

Bolivia (Decreto Supremo 1957),
Costa Rica (1970), Guatemala (1961),
Trinidad y Tobago (Education Act de 1966).

Reformas en los últimos 20 años: 2001–2016

Belice (Education Act 2010),
Santa Catarina, Brasil (2015),
Chile (2016), Colombia (2002), Ecuador (2011),
Honduras (Reglamento de la carrera docente 2013),
México (Ley General del Servicio Profesional Docente, 2013),
Paraguay (Ley 1725 de 2001), Perú (2012).



Fuente: Elaboración propia sobre la base de los estatutos docentes de cada país.

en tres amplios grupos, según la fecha en que se estableció la reglamentación de la carrera: i) antes de la década de 1980, ii) entre los años ochenta y el nuevo milenio y iii) en los últimos veinte años. De acuerdo con Unesco (2015), carreras docentes reglamentadas a través de estatutos o leyes antes de 2001 son, en general, más burocráticas y contienen menos aspectos meritocráticos que las que han sido reformadas más recientemente.

Si bien los procesos ocurridos alrededor de estas reformas recientes están fuera del alcance de esta revisión, vale la pena mencionar algunos aspectos claves que marcaron esos procesos. Por ejemplo, Colombia es el único país cuya reforma determinó la coexistencia de dos estatutos docentes: el Decreto Ley 2277 de 1979 (o antiguo estatuto) y el Decreto Ley 1278 de 2002 (o nuevo estatuto). Como consecuencia de esto hoy en día la fuerza docente está regida por dos reglas de carreras distintas, la primera de las cuales carece de mecanismos que fortalezcan la excelencia docente. Únicamente los docentes contratados por el magisterio después de 2002, y regidos, por lo tanto, por el Estatuto de ese año, son elegidos a través de concursos y están sujetos a evaluaciones docentes con implicaciones en la carrera y formación continua. Para este país, los análisis presentados en este estudio se basan en la reglamentación que surge del estatuto más reciente.⁵

En el caso de Ecuador, como lo detalla Cevallos-Estarellas (2016), entre 2007 y 2013 el país realizó reformas educativas ambiciosas, muchas de ellas enfocadas en los docentes. Entre los cambios introducidos se incluyeron reformas en los programas de formación previa al servicio, la selección, el marco regulatorio, las evaluaciones docentes, la formación continua y las

⁵ En la actualidad, cerca del 50% de los docentes se rige por el antiguo estatuto, el 30% lo hace por el nuevo estatuto y el porcentaje restante corresponde a docentes provisionales.

remuneraciones de los docentes. Todos estos cambios, que buscaban aumentar el prestigio y la calidad de la docencia en el país, modernizaron la carrera y contemplaron aspectos de rendición de cuentas que antes no existían. De igual forma, en Honduras la reglamentación de la Ley Fundamental de Educación (2012) aprobada durante 2014 incorporó cambios importantes en la legislación vinculada con la fuerza docente que comprenden el Reglamento de la Formación Inicial de Docentes, el Reglamento de Formación Permanente de Docentes, el Reglamento de Evaluación Docente y el nuevo Reglamento de la Carrera Docente. Aunque estos nuevos instrumentos todavía no se aplican en el país, que la legislación los contemple es sin dudas un avance.

Perú, por su parte, aprobó en 2012 la Ley de Reforma Magisterial (Ley N° 29944), que modifica las normas de selección de los docentes, su progresión dentro de la carrera y su desarrollo como profesores. La idea fundamental de la nueva reglamentación es que la carrera debe estar regida por la excelencia y el principio de mérito y capacidad. Esto implica que el país está llevando a cabo esfuerzos importantes para establecer un proceso de evaluación capaz de garantizar que ingresen en la profesión y permanezcan en ella los docentes más idóneos, fortalecer los programas de formación y capacitación permanente y mejorar los salarios del gremio.

También para México 2012 fue un punto de inflexión en torno de las políticas docentes. Aunque el país había efectuado reformas importantes previas, como la introducción del Programa de Carrera Magisterial en 1993, solo a partir de 2012 se puso en marcha una reforma del sistema educativo en diversas áreas. En ese sentido, es importante resaltar la creación del Instituto Nacional para la Evaluación de la Educación (INEE), un ente que estaba encargado de regular y supervisar las evaluaciones docentes en el país, y la Ley General del Servicio Profesional Docente, que regulaba un nuevo tipo de carrera meritocrática para los docentes. Sin embargo, estas reformas fueron derogadas en 2019 (en la práctica eliminando a el INEE y la evaluación docente).

Finalmente, vale destacar el caso de Chile, el país donde más recientemente se han llevado a cabo reformas en la reglamentación de la carrera docente. Si bien el primer estatuto data de 1993, en 2016 se establece la Ley N° 20903 que crea el Sistema Nacional de Desarrollo Profesional Docente (SNDPD). El nuevo sistema tiene como objetivo aumentar la valoración de la carrera por medio de una formación inicial de la más alta calidad, asegurar una inserción a la carrera a través de un acompañamiento en los primeros años de ejercicio, elevar la formación continua y una mejora en la remuneración docente de acuerdo con el nivel de desarrollo profesional. Así, la Ley del SNDPD, la Ley de Evaluación Docente de 2004 y la ley 20.501 sobre calidad y equidad de la educación de 2011 promueven una carrera meritocrática para todos los docentes que trabajan en escuelas públicas o subvencionadas en el país.

Las diferencias en los estatutos que definen las reglas que rigen la profesión docente en la región son inmensas. Sin embargo, debido al interés particular de este estudio, se resaltan dos aspectos cruciales relacionados con las evaluaciones docentes. El primero de ellos está ligado al tipo y al número de evaluaciones docentes consignados en cada estatuto y a los cuales, por lo tanto, están sujetos los docentes. El segundo aspecto, que se

analiza en detalle más adelante, tiene que ver con la clase de carrera que determina cada uno de ellos y las posibilidades de ascensos que cada docente tiene.

El análisis de los estatutos permite evidenciar que, salvo contadas excepciones, los sistemas que realizaron reformas recientes han definido carreras más meritocráticas y menos burocráticas que las que existen en sistemas con estatutos que no han sido modificados en más de 30 años. Por otra parte, en los sistemas mencionados en primer lugar la evaluación docente es más completa en términos de diseño (en cuanto a la multiplicidad de instrumentos y a los evaluadores utilizados) e implementación (entendida como la aplicación efectiva de lo establecido en las leyes) y en el uso que se da a los resultados de las evaluaciones cuando efectivamente se hacen. Los detalles de cada tipo de evaluación se presentan en las secciones correspondientes. No obstante, el cuadro 4.1 indica si los

CUADRO 4.1 | Evaluaciones mencionadas de manera explícita en los estatutos de cada sistema educativo

Sistema educativo	Tipo de evaluación docente			
	Concurso docente	Periodo de prueba	Obligatoria periódica	Ascensos horizontales
Prov. Buenos Aires (Argentina)				
Belice				
Bolivia				
Santa Catarina (Brasil)				
Chile				
Colombia				
Costa Rica				
Ecuador				
El Salvador				
Guatemala				
Honduras				
Jamaica				
México				
Panamá				
Paraguay				
Perú				
República Dominicana				
Trinidad y Tobago				
Uruguay				

- Se menciona explícitamente en los estatutos de cada país
- En estos países se otorga una licencia en lugar de concurso
- No existe mención alguna al respecto en los estatutos

Fuente: Elaboración propia sobre la base de Rodríguez y Rivera (2018).

estatutos se refieren de manera explícita a cada una de las cuatro evaluaciones analizadas en este estudio: concursos para entrar a la carrera docente, evaluación de período de prueba, evaluaciones obligatorias periódicas y evaluaciones que definen ascensos horizontales en la carrera. Es decir, establece la situación *de jure* respecto de cada tipo de evaluación para cada sistema educativo.

Al observar el cuadro 4.1 se pueden extraer, al menos, cuatro mensajes claros. El primero es que, sin duda, los países de la región han incluido en la agenda de política docente la discusión acerca de la evaluación en distintas etapas de la carrera y su obligatoriedad. El segundo mensaje es que estos esfuerzos se han concentrado en dos evaluaciones: evaluaciones para el ingreso a la carrera y evaluaciones obligatorias periódicas. Hoy en día la totalidad de los sistemas educativos analizados obliga a que al menos una parte de sus docentes cumpla con requisitos mínimos y presente una evaluación para poder ingresar a la carrera. Asimismo, con excepción de los sistemas de Belice y Santa Catarina (Brasil), todos buscan llevar a cabo procesos de evaluación obligatoria periódica de sus docentes. En tercer lugar, a través de evaluaciones que determinan los ascensos horizontales, se ha buscado permitirle al docente ascender en su carrera en términos de prestigio, recursos económicos y responsabilidades sin necesidad de abandonar el aula de clase. Aunque esta posibilidad y los requisitos meritocráticos relacionados con ella son menos comunes, la evidencia sugiere que es un aspecto que está en la mira de algunas autoridades educativas de la región. Por último, quizás el aspecto en el que menor esfuerzo han hecho los países en términos de evaluación docente es el de las evaluaciones tras un período de prueba.

La simple mención de las evaluaciones y de su obligatoriedad es un avance importante y sienta las bases legales para su implementación. Sin embargo, la efectiva aplicación y la calidad de estas varían de forma notable en la región. Las siguientes secciones detallan las principales características en cada una de estas cuatro evaluaciones.

EVALUACIONES PARA EL INGRESO A LA CARRERA DOCENTE

Tal como describen Elacqua et al. (2018), durante el siglo XX los países latinoamericanos hicieron esfuerzos inmensos por asegurar una mayor cobertura educativa. Por supuesto, para dar respuesta al creciente número de alumnos que entraban al sistema era necesario incrementar también de modo veloz el número de docentes públicos contratados por los gobiernos. Como consecuencia de esto, la mayoría de los países flexibilizó los requisitos de entrada al magisterio y expandió los programas de formación docente sin tener en cuenta criterios de calidad. Estas dos decisiones, unidas a una caída en el salario relativo promedio de los docentes y la apertura de nuevas oportunidades laborales para las mujeres, explican en gran medida el deterioro del prestigio de la profesión en la región.

Como consecuencia directa de esto, la preparación promedio de los que actualmente ejercen la docencia ha disminuido de forma significativa y el número de estudiantes que esperan entrar en el futuro cercano en la profesión ha caído de un modo notable. La baja

calidad disciplinar promedio que existe hoy entre los docentes se observa en los resultados de la prueba TEDS-M, que evaluó los conocimientos de matemáticas de docentes de 17 países, entre ellos Chile, y en la cual los docentes chilenos ocuparon el último lugar. En tanto, evidencia para países como Belice (Näslund-Hadley, Alonzo y Martin, 2013), Costa Rica (OCDE, 2017) y República Dominicana (Caraballo et al., 2014) sugiere que los docentes ni siquiera dominan el contenido disciplinar de lo que enseñan. De manera similar, en pruebas estandarizadas realizadas en Colombia, estudiantes de pedagogía obtienen puntajes inferiores a los del conjunto de próximos graduados de educación superior en el país (Figueroa et al., 2018). Con respecto a las habilidades de los jóvenes que se encuentran próximos a graduarse en la educación secundaria y están interesados en la carrera docente la información existente tampoco es alentadora. Los resultados de las pruebas PISA (Elacqua et al., 2018) y de pruebas nacionales estandarizadas de Chile y Colombia (Alvarado et al., 2012, y García et al., 2014) indican que el nivel de habilidades de jóvenes interesados en la docencia está por debajo del que muestran los interesados en carreras como las ingenierías, las ciencias, la medicina o el derecho.

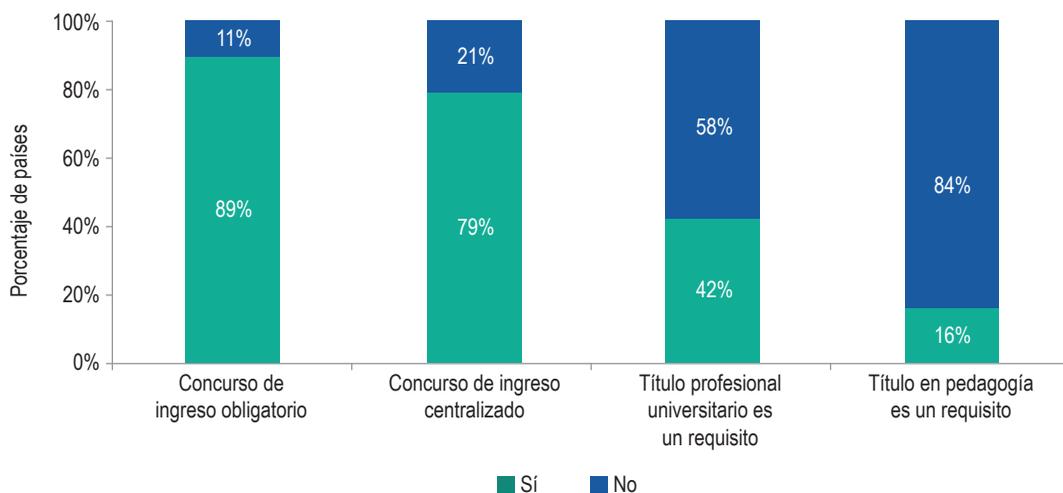
En estas condiciones, los criterios de selección de los candidatos que finalmente son contratados por los gobiernos latinoamericanos para ejercer la profesión docente resultan fundamentales. Solo escogiendo a los individuos mejor preparados y más motivados para esta labor podrá mejorar la calidad de la educación brindada en la región.

La revisión de este aspecto en los 19 sistemas educativos analizados indica que en la actualidad todos ellos cuentan con algún tipo de proceso de selección de candidatos (véase el cuadro 4.1). En dos de ellos, Belice y Jamaica, el Estado otorga una licencia renovable necesaria para poder ejercer la profesión.⁶ En los 17 sistemas restantes la reglamentación vigente obliga a que la mayoría de los candidatos deba participar y aprobar algún tipo de concurso de oposición o de méritos para poder entrar al magisterio. Vale la pena mencionar que hay mucha heterogeneidad en el tipo de concurso, los instrumentos utilizados y el nivel de implementación.

El gráfico 4.4 resume cuatro características básicas de los concursos docentes existentes en los sistemas educativos incluidos en el estudio (los detalles de cada una de estas

6 Toda persona interesada en enseñar en una institución educativa en Belice debe cumplir requerimientos específicos para solicitar al “Chief Education Officer” una licencia que le permita ejercer. Desde un tiempo atrás, el país está en una discusión avanzada que busca homogeneizar y estandarizar los requisitos de ingreso a la carrera docente dentro del magisterio. Se esperaba que en 2018 se introdujera un examen nacional de ingreso que permitiera identificar realmente qué candidatos tenían las calificaciones y la preparación necesarias para llevar a cabo la labor docente en el país. Aunque el país llevó a cabo la primera de estas evaluaciones, sus resultados, sus posibles modificaciones y su aplicación definitiva todavía están en discusión. Por su parte, en Jamaica toda persona graduada de una carrera docente acreditada por el gobierno debe registrarse ante el Ministerio de Educación para quedar certificada como una persona académicamente preparada para ejercer la carrera docente. Una vez cumplido este proceso, la persona obtiene el registro que establece que es un profesional preparado para ejercer la profesión y tiene el permiso del gobierno para poder ejercer como docente en el país.

GRÁFICO 4.4 | Obligatoriedad, centralización y niveles educativos requeridos en los concursos docentes de la región



Fuente: Elaboración propia a partir de los datos de Rodríguez y Rivera (2018).

cuatro características para cada país se encuentran en el cuadro A3 del apéndice). Lo primero que vale la pena resaltar es que a pesar de que todos los estatutos mencionan de manera explícita la necesidad de que el ingreso en la profesión se produzca a través de concursos docentes (o licencias para el caso de Belice y Jamaica) en la realidad esto no sucede. En algunos países, los concursos no se aplican a todos los docentes en el sistema, lo cual deja espacio para una posible corrupción y la pérdida de una oportunidad para seleccionar a los mejores candidatos. Específicamente, los concursos no son obligatorios para todos los docentes en dos sistemas educativos: el de Bolivia y el de Guatemala. En Bolivia, de acuerdo con el artículo 13 del Escalafón Nacional del Servicio de Educación, los docentes normalistas —aquellos que estudian una carrera pedagógica— ingresan a la carrera docente sin necesidad de realizar un examen o de aprobar un concurso de oposición. El concurso es obligatorio para los docentes interinos, es decir para aquellos con otro tipo de profesión en sus estudios de pregrado. En Guatemala, el concurso docente es obligatorio para docentes interesados en enseñar en el nivel de preprimaria o primaria. Sin embargo, los interesados en ser docentes de educación media no deben pasar por un concurso de oposición y son los directores quienes escogen a los docentes.

La segunda característica analizada es el nivel de centralización de los concursos docentes. Como es posible observar en el gráfico 4.4, los concursos son un proceso centralizado en la gran mayoría de los países analizados: independientemente de la región o el nivel educativo en los que quiera enseñar el candidato, todos los concursos están sujetos al mismo proceso. A pesar de que en Belice y Jamaica el gobierno otorga licencias para ejercer como docente y no hay un concurso como tal, fueron catalogados como países donde

el proceso es centralizado porque efectivamente solo una institución puede otorgar dichas licencias en cada país y los requisitos son iguales para todos. Argentina y Brasil, en tanto que tienen una organización gubernamental federal, cuentan con procesos descentralizados con características, instrumentos y exigencias que dependen de lo que establece cada provincia, estado o municipio. De la misma forma, debido al nivel organizacional gubernamental que los caracteriza, Bolivia, Colombia, Costa Rica, Ecuador, El Salvador, Guatemala, Honduras, Panamá, Paraguay, Perú, República Dominicana, Trinidad y Tobago y Uruguay tienen procesos centralizados liderados por el ministerio de Educación de cada país. Cuatro países merecen una aclaración especial en cuanto a la centralización de los concursos. El primero es Chile, porque en ese país la reglamentación de la carrera docente está centralizada de manera detallada en la Ley del Sistema Nacional de Desarrollo Profesional Docente (SNDPD), pero el ingreso de nuevos docentes al sistema es descentralizado. Por el contrario, en México, que tiene un sistema federal, el proceso de los concursos docentes se puede catalogar como centralizado pues el Instituto Nacional para la Evaluación de la Educación (INEE) define los lineamientos y criterios técnicos para la evaluación y la Secretaría de Educación Pública (SEP) y las autoridades educativas y los organismos descentralizados correspondientes los implementan. Asimismo, es importante resaltar el caso de Uruguay, donde las características de la evaluación docente varían de manera significativa según el nivel educativo en el que enseña el docente. En Uruguay existen cinco consejos educativos que están asociados a los distintos niveles y dirigen las políticas educativas en cada uno de ellos de forma independiente.⁷ Esta división dificulta tanto el entendimiento como el manejo del tema docente en el país (INEEd, 2017). En el caso de los concursos, por ejemplo, mientras en el nivel primario se llevan a cabo concursos anuales, la periodicidad de los concursos en el nivel secundario es mucho más espaciada. Además, los dos niveles tienen requisitos e instrumentos diferentes. Por último, en Perú el Concurso de Ingreso a la Carrera Pública Magisterial se realiza cada dos años y se compone de dos etapas. La primera etapa la ejecuta el Ministerio de Educación y consta de un examen, conocido como la Prueba Única Nacional, en el que se evalúan habilidades generales y conocimientos disciplinares y pedagógicos. Quienes aprueban este examen son habilitados para presentarse a la segunda etapa del concurso. Esta nueva etapa se efectúa en cada institución educativa con plazas disponibles y está a cargo de un comité de evaluación, entre cuyos miembros se encuentran el director y el subdirector o coordinador académico. En esta segunda etapa se evalúan la experiencia, la formación docente y las capacidades didácticas y pedagógicas del docente.

En lo referente a los requisitos para participar en los concursos, la mayoría de los países exige que el candidato sea nacional o residente autorizado en el país, que goce de

⁷ Los consejos en Uruguay son: Consejo Directivo Central, Consejo de Educación Inicial y Primaria, Consejo de Educación Secundaria, Consejo de Educación Técnico-Profesional y Consejo de Formación en Educación.

buena salud y que no tenga antecedentes judiciales que le impidan trabajar en una escuela. Los estatutos también detallan el nivel de formación inicial necesario para participar en el concurso. Como se observa en el gráfico 4.4 (la tercera y la cuarta barras resumen los requisitos educativos impuestos por cada país para ingresar al magisterio), solo el 42% de los países exige que el interesado en ser docente público posea un título profesional universitario. En algunos casos el requisito del nivel educativo no está explícito y únicamente se menciona, por ejemplo, la necesidad de tener un título “habilitante” sin dar más detalles al respecto. De hecho, en muchos países, como Brasil, Colombia, Honduras, Trinidad y Tobago y Uruguay, los estatutos docentes establecen de manera clara que los interesados con títulos académicos inferiores a los universitarios pueden presentarse al concurso para acceder, generalmente, a un cargo docente en el nivel primario. Por ejemplo, en varios de estos países se autoriza el ingreso a la profesión de individuos graduados de escuelas normales. Lo fundamental, sin dudas, es que la formación que se dé en estas escuelas les brinde las herramientas pedagógicas y disciplinares necesarias para un correcto desempeño profesional.⁸

Finalmente, al analizar si los estatutos exigen o no un título en pedagogía, se encontró que solo en tres sistemas esto es obligatorio: provincia de Buenos Aires (Argentina), Jamaica y Perú. En los sistemas restantes (el 85%) pueden participar en el concurso de profesionales de otras áreas y si los aprueban, en la mayoría de los casos tienen un tiempo máximo para complementar sus estudios con programas específicos en pedagogía mientras ejercen la profesión.

Para mejorar los procesos de selección docente en la región es preciso contar con concursos obligatorios para todo interesado en la carrera magisterial. No obstante, el punto crucial que ayudará a la selección de candidatos idóneos para la labor dependerá de los instrumentos evaluativos que estos tengan.⁹ El cuadro 4.2 resume los principales componentes que integran los concursos docentes en cada uno de los 17 sistemas analizados donde existen y están reglamentados (por eso se excluyen de este listado Belice y Jamaica). Entre ellos se destacan algunas particularidades. En la provincia de Buenos Aires (Argentina) y Bolivia la reglamentación es clara respecto de que los concursos deben contener exámenes de conocimiento y coloquios. Sin embargo, no existe información acerca de si estos exámenes se aplican efectivamente y qué tipo de conocimientos evalúan cuando se aplican. Debido a que en Chile la realización de los concursos está descentralizada y no

⁸ En varios países aún hay docentes sin título profesional trabajando en escuelas de enseñanza secundaria. Por ejemplo, la revisión encontró que en países como Belice solo el 30% de los docentes en el nivel secundario está entrenado para ser docente. De manera similar, en Uruguay, donde la proporción de docentes con un título universitario pedagógico llega casi al 100% en la enseñanza primaria y es de las más altas en la región, los niveles de titulación de los docentes en la enseñanza secundaria son mucho más bajos. De acuerdo con el informe de la OCDE (2016) cerca del 50% de los docentes en ese nivel educativo no había completado su preparación universitaria según el censo de docentes de 2007.

⁹ El cuadro A4 del apéndice detalla para cada país los componentes de los concursos docentes de acuerdo con lo establecido en la legislación.

CUADRO 4.2 | Número y tipo de instrumentos utilizados en los concursos docentes en la región

Número de instrumentos incluidos	País, provincia o ciudad al que pertenece el sistema educativo	Porcentaje de sistemas educativos
Ningún instrumento/sin información	Chile, Costa Rica, El Salvador, Honduras, Panamá	29%
Un instrumento		
<i>Examen de conocimiento</i>	Bolivia, Santa Catarina (Brasil), Guatemala, México, Uruguay	29%
<i>Entrevista</i>	Trinidad y Tobago	6%
Dos instrumentos		
<i>Examen de conocimiento y entrevista</i>	Provincia de Buenos Aires (Argentina), Colombia, Paraguay, República Dominicana	24%
<i>Examen de conocimiento y clase demostrativa</i>	Ecuador	6%
Tres instrumentos		
<i>Examen de conocimiento, entrevista y clase demostrativa</i>	Perú	6%

Fuente: Elaboración propia a partir de Rodríguez y Rivera (2018).

se detallan los tipos de instrumentos utilizados se optó por dejar el caso como “sin información disponible”. En Costa Rica, El Salvador, Honduras y Panamá se encontró que en la práctica no se utiliza ningún instrumento para el reclutamiento de docentes nuevos en el magisterio. En Costa Rica, El Salvador y Panamá la revisión determinó que la selección de candidatos se basa en el análisis de los títulos académicos, los cursos realizados y la antigüedad. En Honduras la evidencia sugiere que el sistema no tiene en cuenta aspectos meritocráticos (FEREMA, 2015).

Teniendo en cuenta las aclaraciones respectivas, el análisis del cuadro permite dividir los sistemas educativos en tres grandes grupos según el número y el tipo de instrumentos utilizados en los concursos de ingreso al magisterio. En el primer grupo se ubican los sistemas de los que no se tiene información acerca del concurso o los que no utilizan ningún instrumento para la selección de los docentes públicos que entran al magisterio. En la mayoría de los casos solo se revisan los títulos académicos y en algunos casos particulares prima la corrupción al momento de escoger los candidatos. Debido a que la calidad de muchos de los programas de formación inicial en la región se encuentra cuestionada (Elacqua et al., 2018) puede no estar asegurado que los candidatos seleccionados en estos sistemas estén correctamente formados y cuenten con los niveles de conocimiento adecuados.

En el segundo grupo los sistemas educativos contemplan un único instrumento de selección para definir si el candidato es idóneo para entrar a la fuerza docente pública, el cual se basa generalmente en un examen de conocimiento. La inclusión de áreas evaluadas en estos exámenes varía según el país y pueden contar con aspectos disciplinares, pedagógicos y/o psicotécnicos. Solo en el caso de Trinidad y Tobago el ingreso se basa

únicamente en una entrevista. El tercer grupo está compuesto por sistemas educativos que dentro del proceso incluyen dos o más tipos de instrumentos; la mayoría utiliza una entrevista como instrumento adicional y complementario a los exámenes de conocimiento. Los procesos de selección en Ecuador y Perú son únicos en la región. En Ecuador se incluye un examen de conocimiento y una clase demostrativa. Perú es el único caso estudiado que incluye tres instrumentos: examen de conocimiento, entrevista y clase demostrativa.

PROGRAMAS DE INDUCCIÓN PARA DOCENTES PRINCIPIANTES Y EVALUACIONES PARA SUPERAR EL PERÍODO DE PRUEBA

Expertos argumentan que los primeros años del ejercicio docente son críticos (Ávalos, 2016). Por ejemplo, Jensen et al. (2012) indican que los docentes principiantes dedican un menor tiempo efectivo de clase a actividades relacionadas con el aprendizaje, destinan más tiempo para mantener el orden en su salón de clase y reportan menores niveles de autoeficacia en la enseñanza comparados con docentes que tienen más años de experiencia. No sorprende, por lo tanto, que estos primeros años de experiencia impacten también en la efectividad de los docentes, medida a través del valor agregado en los aprendizajes. Diversos estudios de Estados Unidos encuentran que las mejoras en la habilidad docente están concentradas en los tres primeros años (Hanushek et al., 2005; Rivkin, Hanushek y Kain, 2005), a partir de los cuales no hay retornos positivos (Chingos y Peterson, 2011). Para el caso latinoamericano, Araujo et al. (2016) hallan resultados similares en Ecuador, donde los retornos a la experiencia docente se incrementan rápidamente en los primeros tres años, y luego se estabilizan.

Si a esto se suma el hecho de que en varios países de la región los docentes novatos comienzan su experiencia laboral en escuelas que, en promedio, son vulnerables o están en contextos vulnerables debido a la zona geográfica donde se ubican y las características socioeconómicas de sus estudiantes (García et al. 2014), los programas que promueven la adecuada inserción en el sistema resultan fundamentales. Diversos estudios realizados en Estados Unidos sugieren que los docentes más efectivos se concentran en escuelas con estudiantes de mayor nivel socioeconómico o con menor participación de estudiantes pertenecientes a minorías (Steele et al., 2015). En América Latina también hay evidencia de una distribución inequitativa de docentes. Aunque no hay estudios que analicen directamente la efectividad docente, algunas investigaciones revelan que los docentes menos preparados y peor evaluados suelen trabajar en escuelas con estudiantes de menor nivel socioeconómico (NSE) y peores resultados en las pruebas de aprendizaje. Bertoni et al. (2018) analizan la distribución de docentes en escuelas públicas en Chile, Colombia y Perú y encuentran que los docentes menos calificados (con menor experiencia y con contratos temporales) suelen estar concentrados en escuelas con estudiantes de menor NSE, particularmente en zonas rurales. De forma similar, en México los docentes con poca experiencia y menor nivel educativo suelen concentrarse en escuelas rurales y con estudiantes cuyos padres tienen un bajo nivel educativo (Luschei, Chudgar y Rew, 2013).

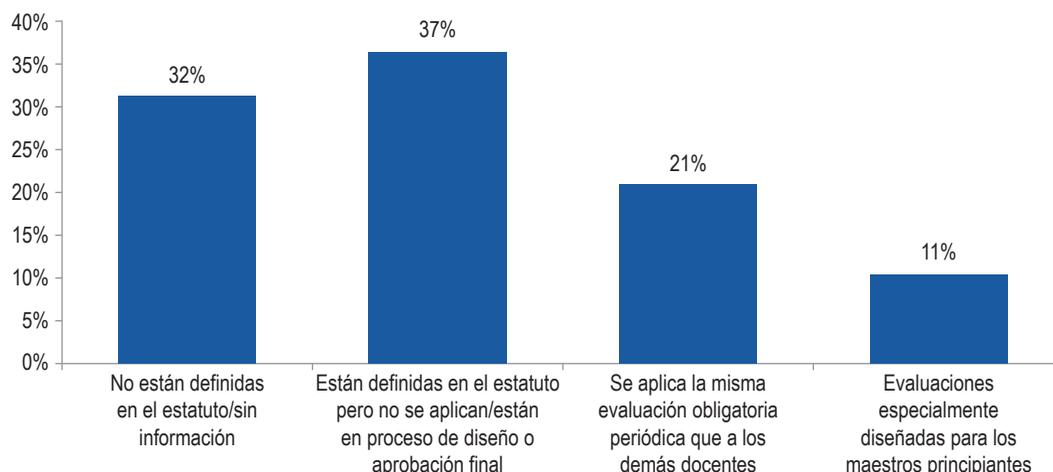
La revisión del tipo de programas de inducción, su estructura y su organización está fuera del alcance del presente estudio. No obstante, la investigación acerca de las evaluaciones efectuadas en los primeros años de ejercicio de la profesión permitió conocer en qué países existe un programa de inducción docente. Así se halló que cerca del 53% de los sistemas educativos analizados cuenta con un programa de inducción impulsado formalmente por el ministerio del área, de modo que, al menos en la teoría, debería beneficiar a la totalidad de los docentes novatos contratados (las principales características de los programas se resumen en el cuadro A5 del apéndice). En muchos casos el programa se basa en tutores o mentores que guían y aconsejan al docente principiante. Como en general los programas son relativamente recientes, ya que muchas veces han sido introducidos por las reformas a los estatutos, la evidencia acerca de su impacto en la región es casi inexistente.

La existencia de un programa de inducción no está perfectamente correlacionada con la existencia de un período de prueba y evaluaciones específicas que determinen o no la permanencia de los docentes principiantes en el sistema, que es el interés primordial de esta sección y un tema de suma importancia para la política docente en general. Como se argumenta en los párrafos anteriores, el inicio de la profesión es un momento fundamental dentro de la carrera docente. Además, es probable que el maestro en cuestión, sus pares y las directivas de los colegios perciban en estos primeros años si el docente principiante tiene la motivación, el conocimiento y la aptitud requeridos en la profesión; cuáles son las debilidades sobre las cuales debería trabajar, y cuáles son sus mayores fortalezas. Evaluaciones de desempeño realizadas en los primeros años del ejercicio de la profesión podrían proporcionar elementos objetivos para esta diferenciación.

Las principales características de las evaluaciones en esta primera etapa de la carrera docente que surgen de la revisión de la reglamentación en los 19 sistemas educativos analizados se presentan en el cuadro A6 del apéndice. Allí se detalla si existe o no un período de prueba al que están sujetos los docentes recién contratados y su tiempo de duración; si están establecidas evaluaciones específicas durante dicho período o al final del mismo, así como si estas evaluaciones determinan la permanencia o no de los docentes en el sistema público. Finalmente, se describen las principales características de las evaluaciones del período de prueba existentes.

El análisis de la información contenida en el cuadro A6 del apéndice se resume en el gráfico 4.5. De ella se deduce que en un tercio de los sistemas educativos que integran el estudio no está definido un período de prueba y por lo tanto no existe una evaluación específica en esta primera etapa de la carrera docente. Vale la pena aclarar un punto a este respecto. Ecuador cuenta con un programa de inducción a docentes principiantes que contempla la evaluación permanente. Sin embargo, estas evaluaciones —que no están definidas en el estatuto— no siguen estándares y no tienen repercusiones en el progreso o la estabilidad de la carrera del docente.

Aunque en los 13 sistemas educativos restantes los estatutos docentes definen claramente un período de prueba de un tiempo determinado, al final del cual el docente debe

GRÁFICO 4.5 | Tipos de evaluaciones del período de prueba a las cuales están sujetos los docentes principiantes

Fuente: Elaboración propia a partir de los datos de Rodríguez y Rivera (2018).

presentar y aprobar algún tipo de evaluación, es necesario resaltar que la situación *de jure* es distinta de la *de facto*. De hecho, en siete de dichos sistemas educativos las evaluaciones para docentes principiantes no se aplican o no se encontró evidencia de que efectivamente se llevan a cabo. Esto último ocurre en Bolivia, donde no existe evidencia de la aplicación de algún período de prueba o de la realización de una evaluación a los docentes no normalistas al cumplir cinco años de ejercicio de la profesión. En Belice, Costa Rica y República Dominicana estas evaluaciones simplemente no se aplican. Por último, en Jamaica, Paraguay, y Trinidad y Tobago, la información recolectada sugiere que las evaluaciones especialmente diseñadas para esta etapa están en proceso de aprobación e implementación final o en etapas de diseño, por lo cual no existe evidencia de que estén siendo aplicadas en su totalidad.¹⁰ Al agregar las primeras dos barras del gráfico 4.5 se explica que cerca de dos tercios de los sistemas educativos de la región no tenían o no implementaban ningún tipo de evaluación para los docentes principiantes.

De acuerdo con la revisión, solo seis de los sistemas educativos de la muestra aplicaban algún tipo de prueba específica a los docentes principiantes: provincia de Buenos Aires (Argentina), Santa Catarina (Brasil), Chile, Colombia, México y Panamá. El cuadro A7 del apéndice detalla quién es el responsable directo de llevar a cabo la evaluación y cuáles son los principales instrumentos utilizados en estos seis sistemas. En cuatro de estos sistemas

¹⁰ En Jamaica, los nuevos formatos de evaluación docente, incluso el de docentes principiantes, debían ser aprobados en septiembre de 2017, fecha en la que también se esperaba el inicio de su implementación.

los docentes principiantes están sujetos a la misma evaluación periódica que se aplica a los demás docentes.¹¹

Únicamente se encontraron dos casos que implementan evaluaciones específicas para los docentes que están iniciando su carrera: el estado de Santa Catarina en Brasil y Colombia. En provincia de Buenos Aires (Argentina), Colombia y Panamá la evaluación de prueba al docente la efectúa el director o superior jerárquico sobre la base de formatos preestablecidos que, por lo tanto, poco pueden diferenciar realmente las habilidades y las debilidades de los docentes. En Santa Catarina (Brasil), Chile y México, por el contrario, la evaluación se basa en múltiples instrumentos y múltiples observadores. En Santa Catarina, la encargada es una comisión evaluadora, un órgano central de la Secretaría de Educación del estado con representación en cada escuela. El órgano central de la secretaría está compuesto por un gerente de área, un coordinador y un servidor de área, mientras que la comisión escolar incluye al director, a un oficial de dirección, a un asistente de educación, a un especialista en asuntos educativos y a un representante de los docentes. En Chile, que aplica la evaluación periódica, las pruebas a los docentes que cumplen dos años en el sistema las realizan evaluadores mediante una multiplicidad de instrumentos. En México, por su parte, los encargados de la evaluación de segundo año son el director, las autoridades educativas locales, la SEP y el INEE. Como menciona el informe del INEE (2018), la evaluación que presentaron los docentes en México al término de su segundo año, que de acuerdo con la reglamentación determina la permanencia o no en el sistema, fue la misma que se aplicó a los docentes en servicio en 2015.

Excepto en Ecuador y Jamaica, la legislación que regula estas evaluaciones en los sistemas educativos analizados establece que los docentes principiantes tienen, en promedio, dos oportunidades para aprobarlas; de lo contrario son retirados del sistema. Solo Jamaica y Uruguay asumen las evaluaciones en esta etapa de la carrera de forma puramente formativa. Aunque la realidad es que el retiro de docentes casi nunca ocurre, incluso en países como México donde las pruebas comprenden una multiplicidad de instrumentos y de evaluadores, y son más detalladas. Por ejemplo, el 99% de los docentes que ingresaron en 2015 y presentaron la evaluación del desempeño al término de su segundo año aprobó el examen con una calificación de “Cumple con la función” (INEE, 2018).

EVALUACIONES OBLIGATORIAS PERIÓDICAS

Luego de que los individuos interesados en la docencia aprueban los procesos de selección, son contratados formalmente por el ministerio, entran en el escalafón docente y superan el período de prueba —en los casos en que esto es un requisito— comienzan su carrera profesional. El resumen de una cuidadosa revisión de la legislación que define y

¹¹ En Chile el estatuto no establece un período de prueba ni una evaluación especial a los docentes principiantes. Sin embargo, a los dos años de ejercicio todos ellos deben presentar las pruebas de la evaluación de desempeño que se detalla en la siguiente sección.

estructura las evaluaciones obligatorias para los docentes que ya han sido nombrados en sus cargos se encuentra en el cuadro A8 del apéndice. Allí se destacan los aspectos más relevantes relacionados con las evaluaciones obligatorias como, por ejemplo, si están dentro de la legislación, su periodicidad, los instrumentos utilizados, las implicaciones que conllevan sus resultados y, por supuesto, si efectivamente se implementan.

Al estudiar la legislación se advierte que en 17 de los sistemas educativos analizados los docentes del magisterio están sujetos a evaluaciones obligatorias periódicas. El gráfico 4.6 resume la presencia en la legislación y la implementación de este tipo de evaluaciones docentes, el tercero que se analiza en el capítulo. Como es posible observar, en la gran mayoría de los sistemas educativos revisados en la región la legislación menciona de manera explícita las evaluaciones obligatorias periódicas a las que están sujetas los docentes nombrados en sus cargos y estas, en mayor o menor medida, efectivamente se implementan. En Costa Rica y El Salvador, a pesar de que las evaluaciones obligatorias están definidas en los estatutos no se implementan. En Jamaica, la reglamentación y el diseño de las evaluaciones periódicas se hallaban en un profundo proceso de reforma en 2017, por lo cual fueron catalogadas de esa forma. Finalmente, los únicos casos en los que los estatutos no mencionan la obligatoriedad de algún tipo de evaluación docente se dan en Belice y en el estado de Santa Catarina, en Brasil. Por tal motivo, los análisis de esta subsección expuestos a continuación excluyen de manera intencional estos dos sistemas.

Las diferencias en el tipo de evaluación obligatoria que se aplica en los países de la región son inmensas, y la primera característica que denota esto es la periodicidad con las que las evaluaciones se efectúan. De acuerdo con la legislación, en 11 sistemas educativos de los 19 sistemas que integran el estudio los docentes están sujetos a evaluaciones obligatorias que deben realizarse con una periodicidad anual: provincia de Buenos Aires (Argentina), Bolivia, Colombia, Costa Rica, El Salvador, Guatemala, Jamaica, Panamá, Paraguay, Trinidad y Tobago, Uruguay. En los seis sistemas educativos restantes (Chile, Ecuador, Honduras, México, Perú y República Dominicana) la legislación establece que las evaluaciones obligatorias deben realizarse cada tres o cuatro años, dependiendo del país en cuestión.¹² Si bien la periodicidad con la que se llevan a cabo las evaluaciones obligatorias

¹² Es necesario aclarar tres puntos al respecto. El primero es que en Honduras la evaluación consta de dos partes: una interna, en la escuela, que debe realizarse cada año y una externa que debe efectuarse cada tres años. De acuerdo con esta última característica se ha clasificado al país en el segundo grupo. El segundo punto es que en 2017 Paraguay empezó una prueba piloto de un nuevo modelo de evaluación docente. Sin embargo, no hay evidencia de que se esté implementando una evaluación detallada de manera sistemática en este momento. Aunque la reglamentación actual establece que la evaluación debe llevarse a cabo de forma anual, si el nuevo modelo comienza a ser implementado de manera censal para los docentes del país quizás su periodicidad cambie. El tercer punto es que la reglamentación en República Dominicana indica que las evaluaciones deben realizarse en un período no mayor de tres años, pero no fija una periodicidad exacta.

GRÁFICO 4.6 | Evaluaciones obligatorias periódicas en la región

■ Existen en la legislación

■ Existen en la legislación, pero hay evidencia de que no se implementan

■ No existen en la legislación

■ Nueva legislación en espera de aprobación

Fuente: Elaboración propia sobre la base de los datos de Rodríguez y Rivera (2018).

Notas: En Paraguay en 2017 se llevó a cabo una prueba piloto de la nueva evaluación docente, que es detallada e incluye una multiplicidad de instrumentos. En la provincia de Buenos Aires (Argentina), Bolivia y Trinidad y Tobago, aunque la legislación menciona la obligatoriedad de estas pruebas, no existe evidencia concreta acerca de si se implementan o no de manera regular.

CUADRO 4.3 | Implementación efectiva y multiplicidad de instrumentos de la evaluación obligatoria de acuerdo con la periodicidad de su aplicación

Periodicidad	Sistemas educativos	Ocorre efectivamente	Múltiples instrumentos
Anual	Provincia de Buenos Aires (Argentina), Bolivia, Colombia, Costa Rica, El Salvador, Guatemala, Jamaica, Panamá, Paraguay, Trinidad y Tobago, Uruguay.	46%	27%
3–5 años	Chile, Ecuador, Honduras, México, Perú, República Dominicana.	100%	100%

Fuente: Elaboración propia a partir de los datos de Rodríguez y Rivera (2018).

Nota: Belice y Santa Catarina (Brasil) no se incluyen en el cuadro porque en su legislación no hay mención alguna de la evaluación obligatoria periódica.

podría parecer un punto menor, lo cierto es que conlleva un sinnúmero de diferencias que incluyen la probabilidad de que dichas evaluaciones se realicen efectivamente, los instrumentos utilizados y las repercusiones que surgen de los resultados, entre otras.

El cuadro 4.3 divide los 17 sistemas educativos analizados que mencionan las evaluaciones obligatorias periódicas en su legislación en dos grupos de acuerdo con la periodicidad en ella definida: anual, y entre tres y cinco años. Su análisis pone en evidencia dos diferencias claras que surgen en las evaluaciones obligatorias de acuerdo con esta división. La primera es que mientras en todos los sistemas que establecen una periodicidad de evaluación obligatoria de entre tres y cinco años las evaluaciones efectivamente se implementan, en los sistemas donde la periodicidad es anual este porcentaje baja a menos de la mitad.¹³

La segunda diferencia tiene que ver con el tipo de evaluación que se implementa. Según la revisión de la literatura especializada, para que las evaluaciones a docentes logren realmente impactar de manera positiva en la práctica pedagógica de los maestros y con ello en la calidad de la educación que reciben los estudiantes es necesario que las evaluaciones cuenten con múltiples instrumentos (MET, 2013). Debido a la complejidad de la labor docente, solo a través de distintos instrumentos es posible tratar de observar las numerosas facetas que debe alcanzar un profesional de la educación. Como queda claro en el cuadro 4.3, mientras todos los sistemas que implementan estas evaluaciones

¹³ En este grupo únicamente fueron clasificados los sistemas educativos en los que fue posible verificar la implementación de una evaluación estandarizada para todos los docentes. Por eso solo entran en el grupo cinco sistemas educativos de los 11 sistemas que tienen una evaluación obligatoria con periodicidad anual: Colombia, Guatemala, Panamá, Paraguay y Uruguay. La revisión permitió concluir que en Costa Rica y El Salvador las evaluaciones no se aplican. En el caso de Jamaica, a septiembre de 2017 los formatos finales de la nueva evaluación debían ser aprobados, por lo cual hasta ese momento aún no se estaban implementando. Para provincia de Buenos Aires (Argentina), Bolivia y Trinidad y Tobago no fue posible encontrar evidencia de que efectivamente las evaluaciones se implementasen de manera continua y estandarizada. Por otro lado, en el caso de Ecuador, a agosto de 2018 todavía no había sido posible la implementación de la totalidad de los instrumentos establecidos en la legislación.

obligatorias con una periodicidad de entre tres y cinco años utilizan múltiples instrumentos, este porcentaje llega solo al 27% en los sistemas donde la periodicidad es anual. Es importante aclarar cuáles son los tres sistemas que integran el grupo con evaluaciones anuales y fueron catalogados como sistemas que utilizan múltiples instrumentos. El primero es Jamaica, que, como se explicó anteriormente, en 2017 se encontraba en un proceso de transformación de la evaluación docente. Aunque los formatos y la legislación en torno de los requisitos de la evaluación docente obligatoria anual estaban ya definidos e incluían el uso de distintos instrumentos, estos aún no se implementaban en el país. El segundo sistema es el de Paraguay, el cual introdujo también en 2017 un nuevo esquema de evaluación docente con múltiples instrumentos cuyo piloto debía comenzar a aplicarse a finales de año. El tercer país es Uruguay del cual se podría decir que, debido a la división administrativa de los distintos niveles educativos a través de los cinco consejos educativos, la evaluación únicamente funciona para los docentes que trabajan en el nivel primario. Si estos aspectos se tuviesen en cuenta, el porcentaje podría ser incluso menor del 27%.

Las evaluaciones que incluyen múltiples instrumentos, y están correctamente diseñadas e implementadas, son costosas en tiempo y dinero. Pensar en aplicarlas de manera censal y anual a todos los docentes de un país se convierte en un objetivo de política educativa inviable. Por eso no sorprenden las enormes diferencias que existen en el uso de múltiples instrumentos entre ambos grupos de sistemas. Más importante aún es el hecho de que esto deja en claro una lección de política educativa: evaluaciones obligatorias censales detalladas, que aporten información valiosa y que efectivamente sean implementadas, no pueden llevarse a cabo anualmente.

La siguiente pregunta que surge en el diseño de política evaluativa es, naturalmente, qué tipo de instrumentos incluir en estas evaluaciones. Como se mencionó en el segundo capítulo, existe una multiplicidad de instrumentos, cada uno de los cuales tiene sus ventajas y desventajas. La totalidad de los sistemas educativos que implementan las evaluaciones obligatorias anualmente y cuentan con un único instrumento utilizan un reporte del director de la escuela. Estos reportes buscan contemplar aspectos relacionados con la gestión académica, la administrativa y la comunitaria, así como el comportamiento de los docentes. Sin embargo, al ser un reporte básico, que no debe ser complementado con evidencia específica para cada aspecto, termina siendo un instrumento que aporta poca información relevante de las habilidades docentes que pueda utilizarse en pro de la mejora profesional de los maestros y de la calidad educativa del país en general. En ese sentido, es interesante analizar la experiencia en los países que implementan evaluaciones obligatorias con múltiples instrumentos en la región. El resumen de estas decisiones y el número total de instrumentos utilizados en cada país de acuerdo con la legislación se encuentra en el cuadro 4.4.

Como es posible observar, el país con el mayor número de tipos de instrumentos mencionados en la legislación es Ecuador (seis tipos de instrumentos), seguido por Honduras y

CUADRO 4.4 | Instrumentos utilizados en países con evaluaciones docentes periódicas obligatorias con múltiples instrumentos

País	Reporte del director o superior jerárquico	Evaluación de pares	Autoevaluación	Portafolio	Observación en clase	Examen	Encuestas a estudiantes y/o padres de familia	Otros
Anual								
Jamaica								
Paraguay								
Uruguay								
3-5 años								
Chile								
Ecuador								
Honduras								
México								
Perú								
Rep. Dominicana								

Fuente: Elaboración propia sobre la base de Rodríguez y Rivera (2018) usando información de la evaluación para cada país.

Nota: En Honduras, el par evaluador realiza observación en clase utilizando la rúbrica Stallings.

Chile (cinco tipos de instrumentos).¹⁴ La mayoría de los países estableció en su legislación el uso de cuatro tipos de instrumentos. Es el caso de Jamaica, México y República Dominicana. Solo tres legislaciones incluyen una cifra menor: las de Paraguay y Perú, tres tipos de instrumentos, y la de Uruguay, dos tipos de instrumentos. De acuerdo con la revisión de la legislación, los dos instrumentos más utilizados en este grupo de países son los reportes del director o superior jerárquico y algún tipo de observación en clase. Solo Paraguay y Perú no involucran las opiniones de los directores en esta etapa; mientras que solo Honduras y México no incluyen una observación de prácticas pedagógicas del docente como instrumento de evaluación. En casi el 67% de los países la legislación incluye como instrumentos la autoevaluación y el análisis de algún portafolio docente. Los países que no contemplan el primer instrumento son Honduras, Perú y Uruguay; mientras que los que no

¹⁴ Es importante hacer una salvedad respecto del número de instrumentos utilizados en la evaluación de desempeño de tres sistemas. En Chile, de acuerdo con los artículos 12 al 17 del Decreto N° 192 del Reglamento sobre evaluación docente, los instrumentos que deben ser utilizados en cada evaluación docente son: i) autoevaluación, ii) entrevista con par evaluador, iii) informes de referencia de terceros, iv) portafolio del docente. Este último instrumento tiene a su vez tres módulos: i) unidad pedagógica, ii) video de clase y iii) trabajo colaborativo. Según lo apreciado en esta revisión, la observación en clase es un instrumento muy relevante en la región. Por eso se optó por tratarlo como un instrumento adicional. Así es que Chile suma cinco instrumentos en lugar de cuatro. En Ecuador y Honduras, los sistemas de evaluación incluyen tanto encuestas a padres de familia como encuestas a estudiantes, por lo que si se consideran estos dos instrumentos independientemente se podría decir que usan un total de siete y seis, respectivamente.

incluyen el análisis de un portafolio son Perú, República Dominicana y Uruguay. Los instrumentos que se mencionan con menor frecuencia en las legislaciones de los países son el examen de conocimientos —existe en la legislación de Ecuador, Honduras y México— y las encuestas a estudiantes y/o padres de familia —contempladas por Ecuador, Honduras y Perú—. Finalmente, solo Chile y Honduras incluyen como instrumento los reportes de pares, mientras que en Perú y República Dominicana se emplean instrumentos específicos para evaluar la gestión del espacio y la responsabilidad (en Perú se utilizan ambos y en República Dominicana se usa solo el último instrumento mencionado).

Para que una evaluación otorgue información certera, válida y confiable se requiere, además de la mención de multiplicidad de instrumentos, su utilización efectiva. Asimismo, hay aspectos relacionados con su diseño —como quiénes son los encargados de llevarla a cabo y qué actores participan en cada etapa— que impactan de manera importante en los resultados. Como en el grupo de países que usan un único instrumento todos utilizan el reporte del director o superior jerárquico este es el actor principal sobre el cual recae la responsabilidad de la evaluación. El punto interesante es, por lo tanto, analizar qué actores intervienen en los países que tienen una evaluación obligatoria con múltiples instrumentos. Una primera aproximación a esta cuestión se presenta en el cuadro 4.5, que muestra qué tipo de evaluadores están involucrados en los procesos de evaluación en cada uno de estos países.

Llama la atención que a pesar de que no todos los países incluyen un reporte del director o del superior jerárquico como instrumento este actor participa de una u otra forma en los procesos de evaluación obligatoria en todos estos países. Específicamente, en Perú el director

CUADRO 4.5 | Actores que participan en la evaluación obligatoria periódica en países con múltiples instrumentos

País	Director/superior inmediato	Par	Propio docente	Expertos calificadoros	Estudiantes/padres
Dos evaluadores					
Uruguay					
Tres evaluadores					
Jamaica					
México					
Paraguay					
Perú					
República Dominicana					
Cuatro o más evaluadores					
Chile					
Ecuador					
Honduras					
Frecuencia evaluador	100,0%	55,6%	77,8%	77,8%	33,3%

Fuente: Elaboración propia sobre la base Rodríguez y Rivera (2018).

forma parte del Comité de Evaluación de Desempeño Docente, conformado también por el subdirector o coordinador del nivel y un profesor del mismo nivel educativo que el profesor por evaluar (o de un nivel superior) perteneciente a otra institución educativa. En Paraguay, el director y un par docente son los encargados de llevar a cabo la observación en clase.

Los segundos actores más comúnmente involucrados en estas evaluaciones con múltiples instrumentos son un experto externo a la institución o escuela y el propio docente. La inclusión de evaluadores externos es, sin dudas, un aspecto positivo que vale la pena resaltar. La ventaja de tener este tipo de evaluadores es que pueden brindarles tranquilidad a los docentes respecto de que, en principio, sus resultados van a estar ausentes de subjetividades por quien calificará su labor en el día a día. Los países en los que la autoridad educativa (ministerios o consejos, en el caso de Uruguay) diseña las evaluaciones, pero asigna personal externo a las instituciones para llevar a cabo parte de ellas han sido catalogados en un primer nivel de independencia. Por ejemplo, en Uruguay la evaluación externa, correspondiente a la observación de clase, está a cargo de individuos conocidos como inspectores, quienes reportan directamente a los consejos respectivos. En Paraguay, el papel de evaluador externo lo cumple un par académico de otra institución educativa. En un segundo nivel de independencia se encuentran Honduras y República Dominicana. En Honduras existe la Dirección General de Currículo y Evaluación, que integra la Secretaría de Educación y se encarga de coordinar todo el proceso de las evaluaciones docentes, desde el diseño de los exámenes de conocimiento hasta el entrenamiento y la designación de los observadores externos. En República Dominicana, el encargado del diseño y la implementación de la última evaluación docente es el Instituto Dominicano de Evaluación e Investigación de la Calidad Educativa, un organismo de carácter técnico adscrito al Ministerio de Educación.¹⁵ En Chile, y Ecuador hay instituciones independientes que se ocupan del diseño de la evaluación y de su ejecución, lo cual le otorga aun mayor independencia al proceso: el Centro de Medición MIDE, de la Escuela de Psicología de la Pontificia Universidad Católica de Chile; el Instituto Nacional de Evaluación Educativa en Ecuador. La existencia de estas instituciones permite que, con distintos niveles, el diseño y la implementación de la evaluación obligatoria tengan la autonomía técnica suficiente para evitar que las decisiones en torno de ella dependan de preferencias o presiones del gobierno o del ministerio de turno.

El 77,8% de los países analizados también incluye como actor al propio docente (véase el cuadro 4.5). Esta categoría incluye, además de la obvia participación del docente como sujeto evaluado, un papel específico como instrumento de autoevaluación de su propia

¹⁵ Como se señaló anteriormente, en Colombia el director es el único actor encargado de llevar a cabo la evaluación periódica obligatoria. Sin embargo, vale la pena mencionar que este país cuenta también con el Instituto Colombiano para la Evaluación de la Educación, una entidad especializada del Ministerio de Educación nacional que se encarga de la realización de los exámenes de conocimiento a estudiantes (examen de estado en grado 11º y las Pruebas Saber 3º, 5º y 9º) y las evaluaciones docentes asociadas a la entrada en la carrera y a las posibilidades de ascenso profesional, esta última descrita más adelante.

práctica pedagógica (Chile, Ecuador, Jamaica, México, Paraguay y República Dominicana) o la presentación de un examen de conocimientos (Ecuador, Honduras, México).

La segunda columna del cuadro 4.5, por su parte, indica que la mitad de los países que usan múltiples instrumentos en las evaluaciones periódicas contempla al par del docente como uno de los evaluadores. En Jamaica y Paraguay (de acuerdo con la legislación que estaba en proceso de aprobación a finales de 2017) el par participa en la observación de clase. En Ecuador, los planes originales le asignaban al par la responsabilidad de coadyuvar en la evaluación del portafolio docente. En Honduras, los pares se involucran tanto en la observación de clase como en la medición de la planificación, el desarrollo de la clase, el manejo del tiempo y la evaluación a los alumnos. Finalmente, en Chile el par es el encargado de realizarle una entrevista al docente para conocer un poco más acerca de su práctica pedagógica y el contexto de su trabajo profesional.

Los actores que se incluyen dentro del proceso de evaluación con una menor frecuencia son los estudiantes y los padres de familia, a través de la respuesta de cuestionarios diseñados exclusivamente para ellos. Esto no llama la atención ya que lograr reunirlos (sobre todo a los padres) y que respondan los cuestionarios seguramente complica la logística de manera importante. Solo Ecuador (donde lo establece la legislación), Honduras y Perú los incluyen de forma específica en el proceso.

Además de escoger el tipo de instrumentos que se utilizarán, el diseño de los procesos de evaluación docente implica la toma de otra decisión importante para aquellos sistemas que incluyen una multiplicidad de instrumentos: la determinación del peso que le asignarán a cada uno de ellos. Los análisis de los resultados del proyecto MET (2013) indican que esta es una elección fundamental que tiene implicaciones en la validez y la confiabilidad del proceso. El cuadro 4.6 presenta el peso que le otorga la legislación a cada uno de los instrumentos utilizados en seis países que asignan los pesos de acuerdo con los instrumentos. Como es posible observar, el instrumento que mayor peso obtiene en promedio (el 43%) es el de observación en clase. Si bien Honduras aparece como el único país que no asigna peso a este instrumento, la realidad es que la evaluación del par incluye la observación en clase al igual que la medición de la planificación (véase cuadro A8 del apéndice). El segundo instrumento al cual los países de la región le otorgan mayor peso es el examen de conocimiento, con un promedio del 33%. Los que siguen en orden de importancia, según el peso asignado, son el portafolio y la evaluación del par. Por último, la autoevaluación, el reporte del director y las encuestas a padres de familia y estudiantes reciben un peso parecido cuando se mira a la región en su conjunto.

Tres países no aparecen referenciados en el cuadro 4.6: Jamaica, México y Uruguay. En México, la ponderación en los documentos oficiales del país solo permite concluir que se les asigna un peso del 60% a los aspectos relacionados con la evidencia y la reflexión de la práctica docente, y el 40% restante se le otorga al puntaje obtenido en los exámenes de conocimiento. En Jamaica, según la legislación, la ponderación se asigna de acuerdo con las dimensiones de los principios guías que rigen la profesión docente en el país y no

CUADRO 4.6 | Peso promedio asignado a cada instrumento en los países que utilizan instrumentos múltiples en las evaluaciones obligatorias periódicas de docentes

País	Reporte director o superior jerárquico		Evaluación de pares			Observación en clase		Padres de familia/ estudiantes	
	Reporte superior jerárquico	Evaluación de pares	Autoevaluación	Portafolio	Observación en clase	Examen	Padres de familia/ estudiantes	Otros	
Paraguay			10%		90%				
Chile	10%	20%	10%	17%	34%			10%	
Ecuador	5%		18%	8%	17%	48%	4%		
Honduras	15%	15%		30%		30%	10%		
Perú					50%		17%	33%	
República Dominicana	15%		10%		45%	20%		10%	
Promedio	11%	17%	13%	19%	43%	33%	9%	16%	

Fuente: Elaboración propia sobre la base de Rodríguez y Rivera (2018).

Nota: En Honduras, el porcentaje asignado a la evaluación de pares (15%) incluye la observación en clase.

directamente a un instrumento en particular.¹⁶ Algo similar ocurre en Uruguay, donde el docente recibe, por parte del instructor, una calificación final que va del 1 al 100 y que se reparte entre su nivel de antigüedad (hasta 20 puntos), su asistencia (hasta 20 puntos) y su aptitud docente (hasta 60 puntos). Esta última categoría se nutre de la calificación que le da el propio inspector y aquella que le otorga el director de su institución educativa.¹⁷

Al comparar los instrumentos y los pesos asignados por los países de la región que implementan evaluaciones docentes obligatorias y periódicas con lo que ocurre en Estados Unidos y en los otros sistemas analizados en el capítulo anterior surgen diferencias importantes. La primera de ellas tiene que ver con el instrumento de observación en clase. Mientras la totalidad de los sistemas educativos analizados en el capítulo 3 incluyen este instrumento dentro de las evaluaciones docentes, en la región solo se implementa en siete de los 19 sistemas estudiados. Como lo muestran Bruns y Luque (2015), los docentes de la región revelan prácticas pedagógicas que se encuentran lejos de los estándares que la literatura internacional especializada considera adecuados, por lo cual las políticas educativas deberían incluir acciones al respecto de forma prioritaria. Sin embargo, la falta de aplicación de instrumentos como la observación en clase implica indefectiblemente que las autoridades educativas ni siquiera tienen la información relevante para comenzar a mejorar dicha problemática.

¹⁶ Los pesos para cada principio guía son: i) conocimiento disciplinar, el 23%; ii) conocimiento pedagógico, el 23%; iii) liderazgo profesional, el 23%; iv) manejo de la diversidad de estudiantes, el 14%; v) continuo mejoramiento de habilidades profesionales, el 9%; vi) relaciones con padres de familia, el 8%.

¹⁷ No se encontró información con respecto al peso que se le otorga al puntaje dado por cada uno de estos evaluadores. Sin embargo, de acuerdo con el informe de OCDE (2016), la calificación final debe aprobarse en una reunión anual de la cual participan tanto inspectores como directores, y donde estos últimos tienen voz, pero no voto.

La segunda diferencia clara radica en que mientras en Estados Unidos la mayoría de los estados y distritos incluye medidas de rendimiento en pruebas estandarizadas y le asigna, en promedio, el segundo peso relativo más alto de todos, ningún país de la región utiliza este instrumento. Como señala la literatura especializada, prescindir de este tipo de instrumento no implica necesariamente un perjuicio. Medidas de rendimiento estudiantil, aunque son útiles, también tienen diversas desventajas como el hecho de que, en general, no están disponibles para todos los docentes, pueden incentivar comportamientos no adecuados por parte de los docentes —reducir la enseñanza a la preparación de un examen o la copia—, están sujetas a problemas de medición si no han sido diseñadas de forma adecuada y su uso implica para el evaluador la necesidad de ser cuidadoso y estar sujeto a *trade-offs* adicionales al momento de definir con qué metodologías calcular las estimaciones de valor agregado.

La cuestión crucial es que la evaluación, con el conjunto de instrumentos seleccionados, logre efectivamente revelar información valiosa respecto de las prácticas de los docentes. Más importante aún es que esa información sea utilizada de forma proactiva para mejorar la calidad de la educación a través de alguno de los objetivos que las evaluaciones deben buscar alcanzar. Como se ha mencionado antes, para conseguir esto es preciso considerar tres aspectos claves: i) diseño e implementación técnica adecuados, ii) resultados que permitan diferenciar y catalogar a los docentes en grupos distintos de acuerdo con sus conocimientos y habilidades disciplinares y pedagógicas y iii) utilizar de manera adecuada estos instrumentos mediante consecuencias en la carrera docente o la implementación y puesta en marcha de programas de desarrollo profesional pertinentes y de alta calidad. En los párrafos siguientes se resaltan puntos específicos de cada uno de estos temas.

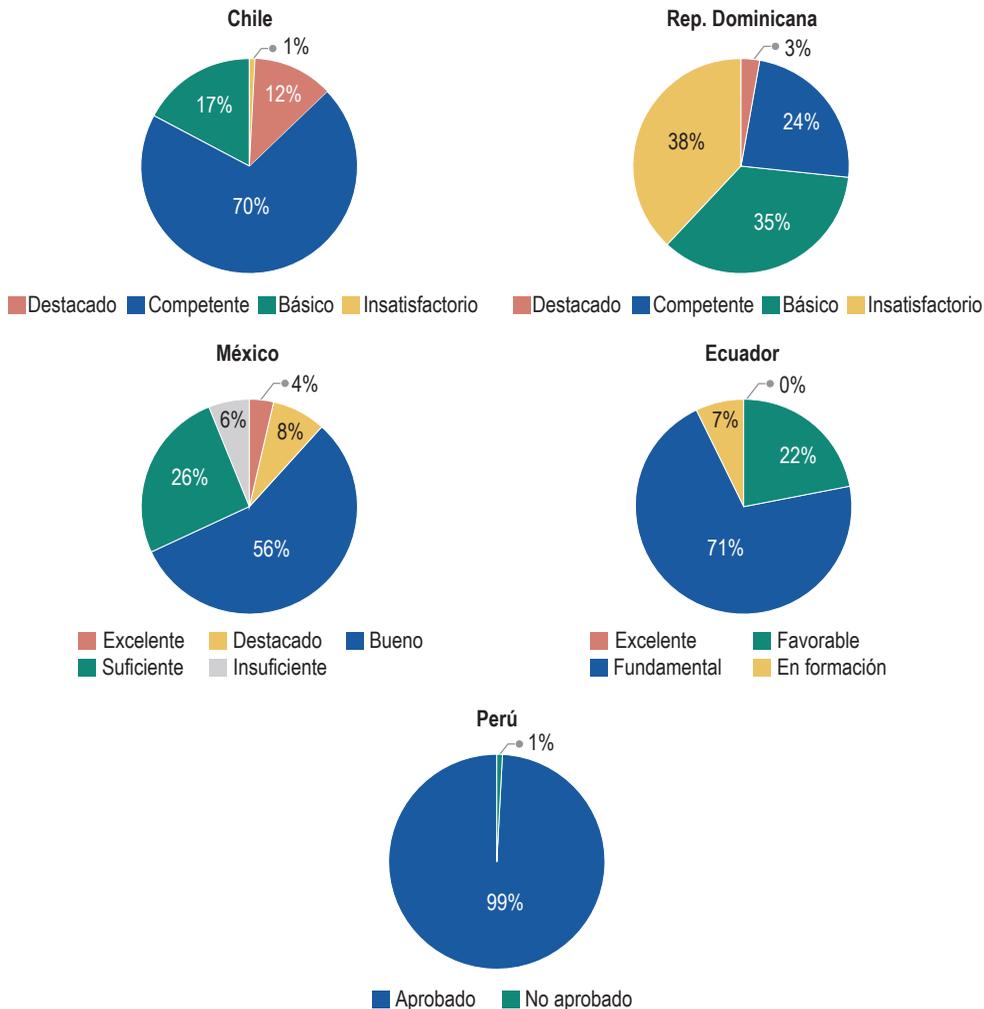
Reseñar los procesos de diseño e implementación de cada instrumento en cada país es un objetivo que está fuera del alcance de este estudio. Sin embargo, la revisión sí permitió identificar cuestiones importantes que vale la pena destacar. La primera de ellas es que no necesariamente todo lo que se establece en la ley se puede cumplir. En el cuadro 4.3 sobre evaluaciones obligatorias se ve con claridad que menos de la mitad de los sistemas educativos cuya legislación contempla una evaluación obligatoria anual la implementa efectivamente. Aunque en el caso de los sistemas que llevan a cabo la evaluación con múltiples instrumentos el porcentaje de implementación llega al 100%, esto no implica que se cumpla a cabalidad lo establecido en la ley. Por ejemplo, en Uruguay, aunque la ley comprende el proceso de observación en clase y fija el número mínimo de observaciones, así como el tratamiento de los resultados que ellas proveen, la realidad está lejos de lo que la legislación reglamenta. A modo de muestra vale decir que en la secundaria el número de inspectores es tan pequeño que muchos de los docentes que enseñan en ese nivel han expresado que nunca fueron evaluados. Asimismo, las diferencias se expresan nítidamente en el reporte que hicieron los propios docentes en la Encuesta Nacional Docente realizada en Uruguay en 2015, donde el 65% de los docentes de la escuela primaria pública manifiesta

haber recibido visitas de un inspector en 2015, mientras que este porcentaje baja al 14% en la secundaria (INEEd, 2017). De manera similar, aunque la legislación de Ecuador menciona que la evaluación docente debe incluir siete instrumentos, la realidad es que a septiembre de 2017 únicamente se había logrado la implementación de manera censal de uno de ellos: el examen de saberes disciplinares.

Por otra parte, aunque se implementen todos los instrumentos mencionados en la ley, su forma de implementación y su calidad pueden variar de forma significativa de un país a otro, como se señaló anteriormente para el caso de los sistemas educativos de alto desempeño. Por ejemplo, en Chile para llevar a cabo la observación en clase (incluida en el módulo 2 del portafolio) se invita al docente a mostrar su trabajo en el aula a través de la grabación de una clase de 40 minutos. Esta grabación, que debe realizar una persona autorizada por Docentemás (responsable de desarrollar las distintas acciones y materiales que requiere el proceso de evaluación) y es gratuita para el docente, luego es evaluada por equipos de docentes de aula en ejercicio en universidades (no son profesores o estudiantes universitarios, como se dice de manera equívoca frecuentemente), bajo la supervisión y el entrenamiento de Docentemás (Docentemás, 2017). Estos docentes de aula tienen formación, y experiencia, en el mismo nivel, la misma asignatura o la misma modalidad de los docentes cuyos portafolios corrigen. Además, si bien estos docentes reciben rúbricas de calificación detallada que buscan asegurar la calidad del trabajo, el 20% de los portafolios es calificado por dos equipos distintos.

De manera similar, es interesante destacar el esfuerzo realizado en República Dominicana donde, con la ayuda de la Organización de Estados Iberoamericanos, el gobierno puso en marcha un proceso censal de observación en clase, que abarca casi 60.000 docentes, con observadores especialmente entrenados para tal fin. A diferencia de Chile, en este país la observación se efectúa en las propias escuelas. Precisamente por la logística, no fue posible hacer la revisión de casos aleatorios o particulares. Por otro lado, en Ecuador, debido a restricciones presupuestales y técnicas, el Instituto Nacional de Evaluación Educativa tomó la decisión de no llevar a cabo la evaluación de prácticas pedagógicas a través de la observación de clase. La opción que permitirá realizar una valoración de práctica de aula de manera viable en el país es la utilización de un instrumento digital que consiste en presentar a los docentes una serie de videos con ejemplos específicos acerca de situaciones cotidianas dentro del aula de clase para que luego los maestros respondan preguntas de opción múltiple mediante la elección de aquellas alternativas que les parezcan más pertinentes para ejecutar en el aula. A fin de 2018, sin embargo, esta etapa de la evaluación aún no había sido aplicada.

Otra de las cuestiones importantes identificadas por la revisión indica que la diferenciación de los docentes de acuerdo con sus habilidades pedagógicas y su conocimiento disciplinar depende, además de los instrumentos y de su implementación, de las categorías o calificaciones que derivan de las evaluaciones. El cuadro A9 del apéndice presenta el análisis para los sistemas en donde la legislación menciona la evaluación obligatoria

GRÁFICO 4.7 | Resultados de la evaluación docente periódica obligatoria en países seleccionados

Fuente: Elaboración propia sobre la base de Rodríguez y Rivera (2018).

y muestra las calificaciones que los docentes pueden obtener en ella. Seis sistemas optaron por asignar un puntaje continuo con niveles específicos que generan, o no, repercusiones en las carreras de los docentes. En los nueve países restantes la legislación establece que la calificación otorgada en el proceso evaluativo es categórica. De hecho, más de la mitad de estos sistemas define cuatro categorías en las que un docente puede quedar clasificado de acuerdo con sus resultados. Solo tres países optaron por un número especial de categorías. Para Perú la evaluación consiste en si el docente aprobó o no; para Colombia consta de tres categorías y finalmente en Costa Rica se

optó por cinco. En el gráfico 4.7 se reportan los resultados más recientes disponibles de la evaluación docente periódica obligatoria para Chile, Ecuador, Honduras, México, Perú y República Dominicana.

El número de categorías importa para la diferenciación docente, pero más categorías no necesariamente implican una mejor diferenciación. Costa Rica, que definió el mayor número de categorías, ni siquiera implementa la evaluación. De la misma manera, para la mayoría de los sistemas que utilizan evaluaciones docentes anuales, los posibles puntajes —basados en el reporte del director— son numéricos y continuos, dando una amplia posibilidad de diferenciación. Sin embargo, se estima que una parte significativa de los docentes de estos sistemas son catalogados con puntajes concentrados en categorías de buenos o sobresalientes. Además, en estos casos una diferenciación amplia en puntajes entre los docentes no resolvería la baja calidad de los reportes de los directores, que impide brindar información valiosa al docente y a las autoridades educativas con respecto a su desempeño laboral.

En el otro extremo se encuentra Perú, donde la evaluación define únicamente dos categorías: aprobó o no aprobó. Sin embargo, al contar con múltiples instrumentos que incluyen la observación en clase, la información recolectada en el proceso puede ser muy valiosa si se usa de manera adecuada. Por ejemplo, la evaluación de docentes de jardines infantiles implementada en 2017 en ese país reportó que el 99,3% de los docentes que participaron en el proceso aprobó la evaluación. Sin embargo, el análisis de los resultados de la rúbrica de observación en clase para estos docentes permite observar la variación significativa existente en aspectos relacionados con el trato del docente a los estudiantes, su sensibilidad ante ellos y la promoción de autonomía y habilidades de lenguaje; e identificar quiénes requieren mejoras específicas en dichos aspectos. Una situación similar ocurre en Chile, donde en 2015 el 82% de los maestros fueron calificados como competentes o destacados, pero que, al analizar los puntajes obtenidos en el instrumento de portafolio el porcentaje bajó al 37%. La información contenida en esta herramienta es precisamente la que permite entender cuáles son las fortalezas de cada docente y en qué aspecto específico es fundamental que mejore. Por último, es útil mencionar también la evidencia recolectada en República Dominicana durante la evaluación docente más reciente. De acuerdo con informes del ministerio, la calificación relacionada con la autoevaluación y el reporte del superior inmediato sugiere que más del 80% de los docentes recibe una calificación superior, pero los resultados de los instrumentos vinculados con la observación en clase y los análisis de planificación y rendimiento profesional son muy distintos: según estos tres instrumentos, solo el 40% de los docentes está en la categoría superior. Además, hallan deficiencias específicas en la probabilidad de que los docentes activen en sus alumnos conocimientos previos, orienten a sus alumnos en actividades dirigidas a la investigación o incluso hagan un uso óptimo del tiempo destinado a la clase.

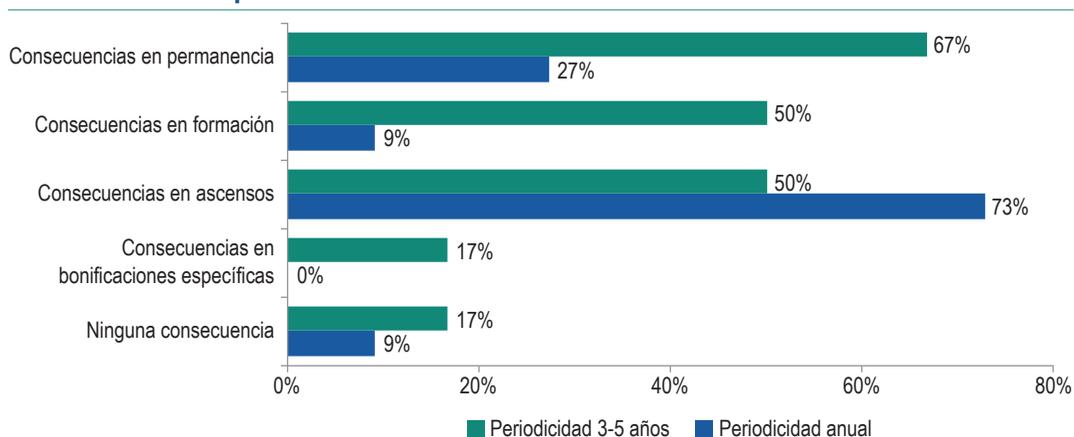
Dentro del estudio de las evaluaciones obligatorias implementadas en la región se analizan también las consecuencias que se derivan de los resultados de cada una, algo que,

por supuesto, está ligado a los objetivos que estas buscan alcanzar. El gráfico 4.8 resume la información recolectada en la revisión, para lo cual divide nuevamente a los países de acuerdo con la periodicidad de las evaluaciones (las consecuencias ligadas a la evaluación obligatoria para cada país se detallan en el cuadro A10 del apéndice). En los sistemas educativos con evaluaciones obligatorias que tienen una periodicidad de entre tres y cinco años las consecuencias más comunes se producen en la permanencia de los docentes (el 67% de los casos) y en la posibilidad de formación y ascenso (el 50% de los casos). En los sistemas educativos con evaluaciones anuales estos porcentajes bajan al 27% y al 9% para permanencia y formación, respectivamente, y la cifra sube al 72% para el caso de los ascensos.

Respecto de la consecuencia de permanencia, por lo general en ambos grupos la legislación establece que el docente debe obtener un puntaje mínimo en la evaluación obligatoria periódica para asegurar su permanencia dentro del magisterio. En caso de que el docente no alcance este puntaje tiene entre una y dos oportunidades adicionales para obtenerlo. En los sistemas con evaluación anual esto ocurre durante el año inmediatamente posterior y en los sistemas con evaluaciones más espaciadas la reglamentación, por lo general, lo fija así también. Si en estas evaluaciones adicionales el docente tampoco alcanza el puntaje mínimo, la legislación estipula que debe ser removido del cargo. Sin embargo, en los países de la región, sin importar el tipo de evaluación periódica que implementen, casi nunca los docentes son alejados de las aulas por un mal desempeño en las pruebas obligatorias periódicas.

En el caso de las implicaciones que tienen los resultados de la evaluación en los procesos de formación docente una vez más lo que dice la ley no es necesariamente lo que

GRÁFICO 4.8 | Consecuencias de los resultados obtenidos en las evaluaciones obligatorias periódicas de acuerdo con la periodicidad de su implementación



Fuente: Elaboración propia sobre la base de los datos de Rodríguez y Rivera (2018).

se aplica en la realidad. Entre los sistemas con evaluaciones anuales solo Paraguay ata los resultados directamente a programas de formación en servicio. Sin embargo, en la práctica, a 2017 Paraguay no había implementado este eje. Entre los sistemas con evaluaciones con múltiples instrumentos, Chile, México y Perú incluyen esta consecuencia de manera específica en la legislación. Sin embargo, aunque se han logrado avances al respecto, la escasa evidencia que existe sugiere que aún falta mucho camino por recorrer en este aspecto para dichos sistemas también. Chile ha venido llevando a cabo procesos de formación ligados a los resultados obtenidos en la evaluación de manera continua. Cortés et al. (2011) analizan algunos aspectos relacionados con los Planes de Superación Profesional (PSP) en Chile y concluyen que los esfuerzos realizados son valiosos, pero aún hay espacio para mejorar. En particular los autores sugieren la necesidad de asegurar la calidad de la formación, que en el momento en que se hizo el estudio era una decisión descentralizada y, por tanto, generaba calidades de capacitaciones muy distintas según el municipio; y señalan la importancia de efectuar las capacitaciones en horarios laborales de los maestros. Asimismo, los autores recalcan que su trabajo no es una evaluación de impacto, y que debido a la trascendencia del tema es fundamental realizar una evaluación de este tipo.

México ha realizado varias rondas de evaluaciones, pero según las propias autoridades educativas aún no se ha logrado integrar los programas de formación dentro del proceso (Schmelkes, 2017). Santibáñez, Rubio y Vázquez (2017) analizan la Estrategia Nacional de Formación Continua y Desarrollo Profesional en ese país y concluyen que los esfuerzos de las autoridades educativas federales y estatales han permitido la mejora de la articulación entre la evaluación y los procesos de formación. Por ejemplo, al ser una formación a distancia —única alternativa posible en la actualidad de acuerdo con las autoridades— es altamente valorado el involucramiento de instituciones de educación superior de alta reputación como encargadas de la formación. De todos modos, el estudio encontró que existe amplio espacio para desarrollar una mayor capacitación previa de los docentes en el conocimiento de herramientas de tecnología de la información, ya que se trata de una modalidad a distancia; y para asegurar la calidad de la formación impartida. De igual forma, el estudio señala que hacen falta mayores recursos invertidos en este eje fundamental. Como en el caso de Cortés et al. (2011), el estudio para México no constituye una evaluación de impacto, lo cual deja en claro el vacío que existe en la región respecto de este factor esencial dentro de la teoría de cambio. Finalmente, en el caso de Perú, aunque la legislación establece que de no aprobar la evaluación el docente debe recibir una capacitación, no fue posible encontrar evidencia de que esto efectivamente esté sucediendo en la práctica.

El gráfico 4.8 evidencia también que para los sistemas con evaluaciones obligatorias anuales es mucho más común utilizar sus resultados como insumo o requisito al momento de analizar los ascensos dentro de la carrera que para los sistemas con evaluaciones cuya periodicidad es de tres a cinco años. Entre los sistemas educativos con evaluaciones

obligatorias anuales, Bolivia, Colombia, Guatemala, Jamaica, Panamá, Paraguay, Trinidad y Tobago y Uruguay cuentan con legislación que les otorga a los resultados la capacidad de definir la posibilidad de ascensos o de funcionar como criterios para dichos procesos. Entre los sistemas educativos que implementan evaluaciones con múltiples instrumentos espaciadas los únicos que no utilizan los resultados de estas como criterios para las posibilidades de ascenso son Ecuador, Honduras y República Dominicana.¹⁸ Las últimas barras del gráfico indican que: i) solo la legislación de República Dominicana ata los resultados de las evaluaciones a bonificaciones monetarias específicas y ii) solo la legislación de Honduras y El Salvador no ata los resultados de la evaluación obligatoria a ningún tipo de consecuencia.¹⁹

Finalmente, vale la pena resaltar la importancia que tienen los esfuerzos que llevan a cabo los ministerios en investigación sobre diseño, implementación e impacto de los sistemas de evaluación docente. Dos ejemplos son los de Chile y México. En Chile, diversos tipos de investigaciones han buscado responder tres preguntas: i) cómo funcionan los distintos instrumentos y cómo es posible mejorarlos, ii) qué percepción tienen los docentes sobre el sistema de evaluación y los instrumentos utilizados y, iii) la pertinencia de las evaluaciones y los instrumentos. En cuanto al primer tema, Sun (2017) menciona al menos dos mejoras recientes cuyos impactos en los resultados de las evaluaciones son evidentes: i) desde 2010 el director está obligado a fundamentar el nivel de destacado otorgado a sus docentes, lo cual determinó que mientras en 2008 y 2009 cerca del 30% de los maestros era calificado como destacado por los directores, en adelante esta proporción cayó a casi la mitad y en promedio solo un 15% de los maestros obtiene esta calificación; ii) desde 2014 la evaluación de pares debe contemplar la observación de prácticas reales, lo cual implicó el dictado de capacitaciones con rúbricas holísticas específicas que debían tener en cuenta para esto. A partir de este cambio las calificaciones dadas por los pares evaluadores cambiaron significativamente en la cola superior, pasando de casi el 30% de maestros catalogados como destacados en 2012 y 2013 al 15% en 2014 y 2015. En un segundo tipo de estudios se ha indagado acerca de la percepción que tienen los docentes respecto del sistema de evaluación y la utilidad y la pertinencia de los instrumentos utilizados. Para ello, las evaluaciones han incluido cuestionarios complementarios opcionales que han preguntado a los docentes por la semejanza entre el portafolio y sus tareas cotidianas o su opinión específica de dicho instrumento. Las investigaciones evidencian que el 72% de los maestros están de acuerdo con que las tareas del portafolio son parecidas

¹⁸ En Ecuador, la reglamentación no es clara y tampoco existe evidencia de que efectivamente estos ascensos se estén llevando a cabo. En Honduras, los criterios que se tienen en cuenta para los ascensos horizontales se basan en el tiempo de servicio y en el nivel de educación. Finalmente, en República Dominicana aún no existe la posibilidad de ascensos horizontales.

¹⁹ Como se mencionó anteriormente, la legislación de El Salvador establece que las evaluaciones deben ser anuales, pero no detalla ninguna característica que estas deban cumplir. Más aún, en la práctica las evaluaciones obligatorias no se realizan en el país.

a las que realizan en su cotidianidad como docentes. De manera similar, el 86% de los docentes creen que realizar el portafolio es útil o muy útil para mejorar la reflexión de sus prácticas pedagógicas (Docentemás, 2017). El tercer tipo de investigaciones está relacionado con la pertinencia de las evaluaciones y los instrumentos. Como se mencionó en el capítulo 2, estudios como el de Taut et al. (2016) han encontrado correlaciones entre la calificación obtenida por los maestros en el portafolio y medidas de valor agregado de los estudiantes.

La segunda experiencia es la de México, que ha investigado el funcionamiento de sus sistemas de evaluación obligatoria y los impactos que ha traído la puesta en marcha de estos sistemas en el progreso en la carrera docente. Por ejemplo, después de la evaluación de desempeño implementada en 2015, el INEE, con la colaboración de la Oficina Regional de Educación de la UNESCO para América Latina y el Caribe (OREALC/UNESCO), llevó a cabo la evaluación de su implementación, que incluía, entre otros aspectos, una encuesta de satisfacción a 10.000 docentes y grupos focales con cerca de 410 docentes. Los resultados de estas investigaciones impulsaron una reforma en el modelo de evaluación obligatoria a partir de 2017. De manera complementaria, México cuenta con investigaciones que buscaban determinar el impacto de aspectos específicos de algunas reformas como la de Estrada (2019) y la de Santibáñez et al. (2006), referenciadas en el capítulo 2, que evaluaron las pruebas de entrada a la carrera y la introducción del Programa de Carrera Magisterial, respectivamente.

Aunque todos estos esfuerzos son valiosos, resulta claro que hacen falta más investigaciones en la región que analicen la validez y la consistencia de los resultados que estas nuevas evaluaciones arrojan, así como los impactos en los docentes y los propios alumnos.

EVALUACIONES PARA ASCENSOS HORIZONTALES

El último tipo de evaluación analizado en el estudio es el que determina si un docente tiene las habilidades para ascender en su carrera. La existencia de estas evaluaciones depende, primero, de la estructura de la carrera docente definida en los estatutos o la legislación de cada país. Por eso, la revisión de esta última evaluación debe comenzar con el análisis de los tipos de carreras a las que entran a trabajar los docentes en la región.

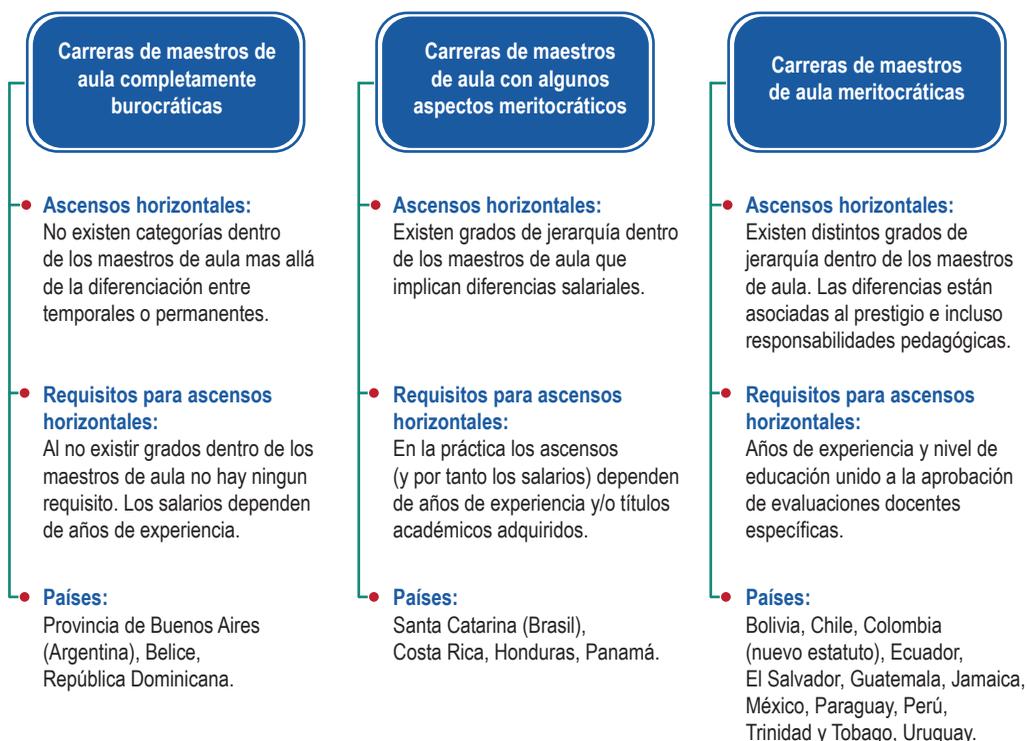
Las diferencias en el tipo de carrera que tienen los docentes en los 19 sistemas educativos analizados son evidentes cuando se analizan sus posibilidades de ascensos y los requisitos que deben cumplir para alcanzarlos. Para entenderlas, es útil dividir los ascensos profesionales en dos clases: verticales y horizontales. Los ascensos verticales son los que obligan a los docentes a dejar las aulas para acceder a trabajos más administrativos y no necesariamente relacionados de forma directa con la docencia. Por ejemplo, los docentes pueden escoger ascender a cargos administrativos dentro de las escuelas o fuera de ellas, como subdirectores o directores, o a cargos dentro del ministerio o de las secretarías de Educación a nivel regional o nacional. Los estatutos de los 19 sistemas educativos

analizados ofrecen esta posibilidad a sus docentes, quienes en general pueden alcanzar estos cargos a través de concursos especialmente creados para ese fin.

Los ascensos horizontales, por su parte, son aquellos que les permiten a los docentes incrementar el prestigio, las funciones, las responsabilidades y el salario sin la necesidad de abandonar las aulas de clase. A diferencia de lo que ocurre con los ascensos verticales, no todos los docentes en la región pueden ascender de manera horizontal en su carrera. De acuerdo con el análisis de los estatutos docentes en este aspecto particular es posible dividir los sistemas educativos en tres grupos, reflejados en el gráfico 4.9, según las posibilidades de ascensos horizontales y los requisitos que deben cumplir los docentes para obtenerlos.

El primer grupo está conformado por los sistemas educativos cuyos estatutos docentes ni siquiera reglamentan categorías diferenciadas de los docentes, al margen de distinguirlos entre los que están contratados de forma permanente y los que no lo están. Se las llama carreras completamente burocráticas porque su reglamentación no les otorga a los

GRÁFICO 4.9 | Posibilidades de ascensos horizontales en las carreras docentes de la región y sus requisitos, según los estatutos docentes



docentes la posibilidad de crecer en términos profesionales. Belice es el caso extremo, ya que allí no existe una carrera definida. Le siguen República Dominicana, donde el estatuto solo define el cargo de educador, y provincia de Buenos Aires (Argentina), donde para docentes de aula solo existen las categorías de docente y docente especial. Lo único que diferencia a los docentes en estos tres sistemas es la cantidad de años que tienen de experiencia enseñando en el sector, lo cual, a su vez, define los salarios que devengan. Vale la pena mencionar que en 2016 República Dominicana diseñó una Guía de Certificación y Desarrollo de la Carrera Docente a fin de catalogar a cada docente en cinco categorías que van de excelente a insuficiente, de acuerdo con sus habilidades pedagógicas, y a su vez atar estas categorías a bonificaciones y reconocimientos especiales. Sin embargo, esta certificación aún no se está aplicando en el país, razón por la cual fue catalogado en este primer grupo.

En el segundo grupo se encuentran los sistemas educativos cuyos estatutos definen de manera explícita distintos niveles o grados de jerarquía para docentes que quieren permanecer en las aulas de clase. Sin embargo, estos ascensos no son meritocráticos y dependen principalmente de requisitos asociados a los años de experiencia en una categoría específica y/o al nivel de educación que tiene el docente. Dentro de este grupo se ubican Santa Catarina (Brasil), Costa Rica, Honduras y Panamá. Aunque en Brasil hay ejemplos de evaluaciones para ascensos, como la introducción de la carrera acelerada en San Pablo, en muchos de los estados y municipios del país, y en particular en Santa Catarina, estado que se analiza en este estudio, el estatuto solo menciona como requisito los títulos académicos.

El tercer grupo está conformado por los países cuyos estatutos ofrecen a sus docentes carreras de carácter meritocrático. Además de permitirles ascender en su profesión con mayores niveles salariales sin abandonar las aulas de clase, en general estos ascensos se asocian también a mayores niveles de prestigio dentro del magisterio por su carácter meritocrático. En algunos casos, las mayores categorías dentro de la carrera les dan a los docentes la posibilidad de ejercer en funciones de líderes o guías pedagógicos como tutores o mentores de sus pares. En estos países la legislación incluye como requisito que los docentes aprueben algún tipo de evaluación y obtengan niveles educativos y de experiencia más altos para lograrlo.

Existe una diferencia adicional entre los países catalogados en el tercer grupo, relacionada con el número de categorías y, por lo tanto, de ascensos horizontales que están reglamentados en cada uno de los estatutos (el cuadro A11 del apéndice presenta las categorías definidas por cada legislación). Mientras que en la mayoría de los países que conforman este grupo las escalas son numéricas o se designan por letras, en Chile y Jamaica los ascensos horizontales se asocian a categorías con una denominación ligada claramente a la experiencia y el prestigio del docente. En Chile hay cinco categorías, que el estatuto llama “tramos”: inicial, temprano, avanzado, experto I y experto II. Cumplir los requisitos para ascender en los primeros tres tramos es obligatorio para todos los docentes en el

país, en tanto que ascender a los últimos dos tramos es un proceso voluntario. Además de completar los años de experiencia requeridos, los docentes expertos deben obtener los mejores resultados en los dos instrumentos de la evaluación de desempeño: el portafolio y la evaluación de conocimientos específicos y pedagógicos. En Jamaica existen tres categorías para docentes de aula: principiante, efectivo y *master*. Mientras ascender de principiante a efectivo es obligatorio si se quiere continuar en la carrera, ascender a la categoría de *master* es opcional.

El detalle de la carrera, las posibilidades de ascenso y los requisitos para alcanzarlos se presentan en el cuadro A12 del apéndice para todos los sistemas. Por supuesto, solo es posible e interesante analizar los requisitos exigidos para ascender para los países catalogados en el tercer grupo del gráfico 4.9. El cuadro 4.7 reseña los instrumentos que, de acuerdo con lo establecido en la legislación, deben ser utilizados al momento de evaluar los ascensos horizontales en esos 12 países, al margen de los requisitos vinculados con el tiempo de servicio o la experiencia exigidos en cada sistema. Llama la atención que cerca del 70% de los países incluye como requisito un componente relacionado con la evaluación obligatoria periódica descrita en la sección anterior, lo cual lo convierte en el insumo más

CUADRO 4.7 | Instrumentos definidos para las evaluaciones de ascenso horizontales en los países donde esta posibilidad existe y la legislación incluye aspectos meritocráticos

País	Promedio/ puntaje o componente evaluación obligatoria	Auto- evaluación	Observación en clase	Examen	Investigación	Encuestas a estudiantes o padres de familia	Matriz de desempeño global	Entrevista
Bolivia								
Chile								
Colombia								
Ecuador								
El Salvador								
Guatemala								
Jamaica								
México								
Paraguay								
Perú								
Trinidad y Tobago								
Uruguay								
Porcentaje	67%	8%	8%	58%	8%	8%	8%	8%

Fuente: Elaboración propia sobre la base de los datos de Rodríguez y Rivera (2018).

común dentro de las evaluaciones de ascenso en la región. En segundo lugar, se ubican los exámenes específicos diseñados para tal fin que evalúan conocimientos disciplinares y pedagógicos. Con excepción de estos dos instrumentos, los requisitos exigidos a los docentes son muy heterogéneos y no existe un patrón claro en su utilización.

Como sucede con las evaluaciones anteriores, lo que se define en la ley no es necesariamente lo que ocurre en la práctica. En la realidad muchos ascensos no dependen de ningún tipo de evaluación que permita evidenciar de forma efectiva las habilidades de los docentes y terminan siendo burocráticos. En Bolivia, la legislación detalla los requisitos que es preciso cumplir para acceder a un ascenso, pero no existe evidencia que sugiera que efectivamente se aplican, algo que, como se ha visto, ocurre también con las demás evaluaciones. En El Salvador la revisión halló evidencias de que los docentes no permiten que se lleve a cabo ninguna evaluación, por lo cual, nuevamente, los ascensos terminan basándose en los niveles de educación y experiencia. Tanto en Guatemala como en Uruguay para determinar los ascensos la ley exige el uso de los puntajes de evaluaciones obligatorias periódicas, pero como quedó claro en el análisis de la sección anterior estas en realidad terminan dependiendo de un único instrumento diligenciado por los superiores jerárquicos que no otorga evidencia detallada de las habilidades docentes. Para Trinidad y Tobago tampoco se halló evidencia de que la entrevista exigida para alcanzar un ascenso esté estructurada de forma adecuada y deba ser complementada con evidencia de la labor docente, por lo cual no es de esperar que otorgue información certera de las prácticas docentes en el país. En este contexto se puede decir que solo en siete de los países analizados en esta investigación los ascensos dependen de evaluaciones específicas que permiten obtener evidencias de las habilidades pedagógicas y disciplinares de los docentes: Chile, Colombia, Ecuador, Jamaica, México, Paraguay y Perú. El cuadro A13 del apéndice detalla que instrumentos se incluyen en las evaluaciones en este grupo de países.

El análisis del cuadro 4.7 y del cuadro A13 del apéndice revela tres aspectos interesantes. El primero es que sistemas educativos que implementan evaluaciones obligatorias periódicas detalladas utilizan los resultados o los componentes específicos de estas evaluaciones al considerar las posibilidades de ascenso de su cuerpo docente. Resulta lógico que luego de invertir recursos importantes en tiempo y dinero en estas evaluaciones obligatorias estos países aprovechen la información recolectada sobre las habilidades docentes. De hecho, en el grupo de sistemas educativos que implementan evaluaciones obligatorias con múltiples instrumentos únicamente Honduras y República Dominicana no utilizan sus resultados en los procesos de ascenso horizontal. En Honduras esto se debe a que la legislación solo establece como requisitos para el progreso en la carrera la experiencia y la educación. En República Dominicana la posibilidad de ascenso horizontal no está definida en el estatuto.

El segundo aspecto relevante tiene que ver con que en la mayoría de estos siete países: Chile, Colombia, Ecuador, Jamaica, México, Paraguay y Perú, donde se incluyen

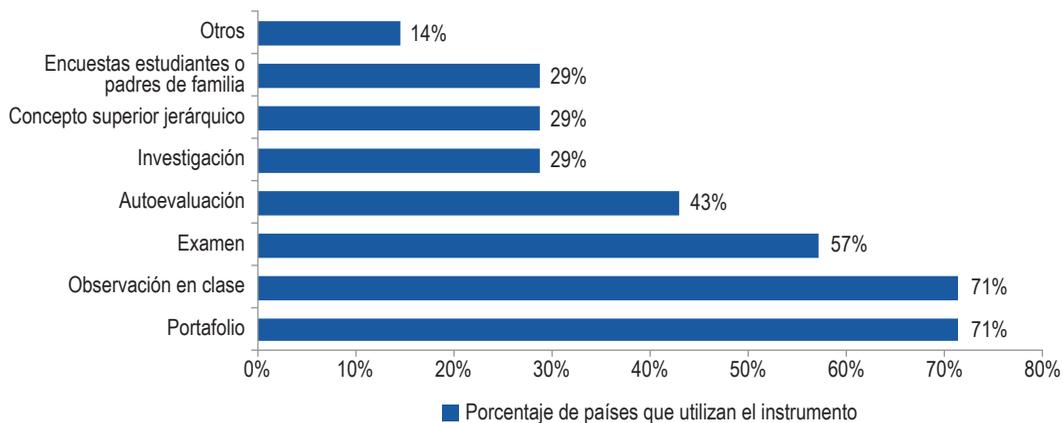
instrumentos adicionales para permitirles a sus docentes ascender. Para Colombia y Jamaica las evaluaciones de ascenso abarcan múltiples instrumentos adicionales, no incluidos en las evaluaciones periódicas obligatorias que se detallan en el siguiente párrafo. Por su parte, Chile y Perú complementan la información de la evaluación obligatoria con exámenes estandarizados adicionales que los docentes interesados en ascender deben presentar. En México, los docentes además de obtener resultados de “Destacado” o “Excelente” en la evaluación de desempeño, debían obtener un resultado “Sobresaliente” o “Relevante” en una evaluación adicional cuyas características debía determinar el INEE. Si bien en Ecuador la reglamentación establece que los docentes deben presentar exámenes adicionales, las reformas que están en proceso de implementación tendrán en cuenta las calificaciones obtenidas en el examen aplicado a los docentes de manera censal en el marco de la evaluación obligatoria en 2016. En Paraguay el estatuto exige que los docentes hayan realizado una investigación académica. Sin embargo, no hay evidencia de que este requisito efectivamente se dé en la práctica y, si ocurre, no hay información acerca de la calidad requerida para estas investigaciones.

El tercer aspecto, mencionado anteriormente, está ligado a que dentro de este grupo de países Colombia y Jamaica son los que tienen diseñado el mayor número de procesos específicos adicionales para esta evaluación asegurando el uso de múltiples instrumentos. Como en Colombia la evaluación periódica se basa en un único instrumento que no otorga realmente información detallada de las habilidades profesionales del docente resulta necesaria la inclusión de aspectos como la observación en clase, la autoevaluación y las encuestas a estudiantes. En ese sentido, es útil resaltar dos puntos: i) el nuevo modelo de evaluación para ascensos, conocido como Evaluación de Carácter Diagnóstico-Formativo para docentes (ECDF), surgió después de un duro proceso de negociación con el sindicato de maestros, el cual abogaba por la modificación de esta prueba que antes se basaba en un examen estandarizado de conocimientos finalmente eliminado (Figuerola et al., 2018) y ii) al igual que para Chile, el instrumento de observación en clase se basa en un video que realiza el propio docente cumpliendo con detalles específicos respecto de su calidad y sus contenidos.²⁰ Además, de acuerdo con lo que indican los instructivos, este video es evaluado por dos pares distintos: uno del contexto regional del docente y otro del contexto nacional. En Jamaica, como se detalló anteriormente, la legislación sobre las evaluaciones periódicas es nueva y, por lo tanto, tampoco existe información previa en este aspecto.

Sobre esta base el gráfico 4.10 resume el tipo de instrumentos que este grupo de siete países tiene en cuenta al momento de llevar a cabo los procesos de evaluación de ascenso.

²⁰ Por ejemplo, cada video debe estar enfocado en la práctica educativa, pedagógica, didáctica y de aula, en una de las clases que imparte el docente. El desarrollo de la clase debe incluir tres momentos: 1) contextualización, ii) desarrollo de la clase, 3) reflexión final (información extraída de la página Maestro 2025 en 2017).

GRÁFICO 4.10 | Frecuencia en el uso de instrumentos en evaluaciones de ascensos horizontales meritocráticos en las carreras docentes de la región



Fuente: Elaboración propia sobre la base de los datos de Rodríguez y Rivera (2018).

Resulta clara la importancia que les otorgan a las prácticas pedagógicas, la planificación y los conocimientos disciplinares pues son los aspectos más evaluados al definir los movimientos horizontales en la carrera. En casi la mitad de los países se incluye un proceso de autoevaluación y reflexión del docente y en menor proporción algunos contemplan la percepción que tienen los superiores, los estudiantes y los padres de familia o la capacidad de hacer investigación del propio docente como criterios de evaluación.

Al existir múltiples instrumentos, surge la pregunta respecto del peso que las autoridades educativas le asignan a cada uno de ellos. A diferencia del caso de las evaluaciones obligatorias periódicas, no es posible resumir la información de manera que cada instrumento obtenga un peso específico ya que, como se mencionó antes, muchos de los países que integran este grupo utilizan como insumo información previamente recolectada en las evaluaciones obligatorias. Por ejemplo, en Chile, el nivel en el que queda un docente en el escalafón depende tanto del puntaje obtenido en el instrumento del portafolio (que en este informe incluye la observación en clase y la planeación) como del examen de conocimiento.²¹ En México los ascensos dependen exclusivamente de la calificación final que el docente obtuvo en la evaluación obligatoria y del examen adicional que debía ser determinado por el INEE, mientras que en Ecuador se basan en el puntaje logrado en el examen de conocimientos disciplinares. Solo en Colombia y Paraguay es posible asignarles una ponderación específica a todos los instrumentos. En Colombia, la observación en clase pesa un 80%, los puntajes obtenidos en la evaluación obligatoria de los dos años previos pesan

²¹ El gráfico A3 del apéndice muestra las combinaciones de puntajes que deben obtener los docentes en Chile para ascender a cada una de las categorías.

un 7,5% y las encuestas a estudiantes y la autoevaluación del docente, el 12,5% restante. En Paraguay, los pesos se reparten de la siguiente forma: el 30% para la experiencia y la certificación profesional, el 50% para el puntaje en la evaluación de desempeño y el 20% para la investigación académica.²²

Finalmente, es interesante resaltar que los ascensos no son la única consecuencia atada a los resultados de esta evaluación. En Chile, por ejemplo, si los docentes no aprueban la evaluación en dos ocasiones consecutivas y, por tanto, no pasan de tramo inicial o temprano son retirados del cargo. El docente que ya estaba en el tramo temprano puede volver al tramo inicial dos años después. Sin embargo, si no avanza al nivel avanzado es retirado del cargo de manera permanente. Por otra parte, como se explicó en la sección anterior, en general los programas de formación en servicio están atados a los resultados de las evaluaciones periódicas. Colombia es el único país que ata los resultados de la evaluación docente de ascenso a programas de formación en servicio. Todos aquellos que no logran aprobar la evaluación de ascenso horizontal deben llevar a cabo un programa de formación que les permita superar las debilidades encontradas por la evaluación. Para ello distintas facultades de Educación en universidades a nivel nacional crearon cursos enfocados en trabajar los criterios de la matriz de la ECDF a fin de mejorar las prácticas de aula de aquellos maestros que no lograron superar las evaluaciones. Sin embargo, no existen reportes que den información respecto de su eficacia y confirmen si las prácticas pedagógicas de los docentes fueron efectivamente superadas.

Como en el caso de la evaluación periódica obligatoria, se necesitan investigaciones que permitan comprobar si los instrumentos escogidos para permitir o no ascensos horizontales en las carreras docentes en la región son efectivos y conocer los impactos que estos han traído en la calidad de la educación. De hecho, la evidencia relacionada con estas evaluaciones es aún muy escasa y la revisión no permitió encontrar ningún estudio que respondiera esta pregunta. Probablemente, el poco tiempo transcurrido desde que países como Chile, Colombia, Ecuador, México y Perú aplicaron las reformas a la carrera e introdujeron los nuevos ascensos horizontales explica esta ausencia de análisis.²³

²² En el caso de Jamaica la legislación no es clara y no permite asignar un peso específico a los instrumentos que se utilizan al momento de escoger a los maestros.

²³ La única excepción es la evaluación de impacto de Santibáñez et al. (2006) relacionada con el Programa de Carrera Magisterial anteriormente mencionado. Sin embargo, este programa fue eliminado y reemplazado en las reformas recientes por la Ley de Servicio Profesional Docente.



CAPÍTULO 5

CINCO CLAVES PARA UNA EVALUACIÓN DOCENTE EXITOSA

A pesar de los indudables avances que se han hecho en los últimos años en términos de cobertura educativa en América Latina y el Caribe, la región necesita con urgencia implementar políticas que le permitan mejorar la calidad de sus sistemas educativos. Una de las estrategias más importantes para cumplir este objetivo es mejorar el estado de la profesión docente, lo cual solo se logrará a través de acciones sistémicas que impacten en la calidad y relevancia de la formación inicial y en servicio; un marco regulatorio meritocrático claro que permita la selección de candidatos a la profesión talentosos y que ofrezca la oportunidad de progresar en la carrera, y salarios y beneficios atractivos; y **un sistema de evaluación que promueva el continuo mejoramiento de la fuerza docente.**

Es necesario que los países de América Latina y el Caribe, comiencen o continúen implementando y mejorando sus sistemas de evaluación docente. Deben desarrollar sistemas que sean válidos y confiables, que utilicen una multiplicidad de instrumentos, con evaluadores que procuren entender en detalle distintos aspectos de los procesos de enseñanza, y que generen resultados útiles para la toma de decisiones de política.

Este estudio analiza los aprendizajes relacionados con la estructura y los instrumentos que las evaluaciones deberían tener, sus posibles impactos, y las características fundamentales de las evaluaciones en algunos sistemas educativos que se caracterizan por su alto desempeño educativo. Asimismo, presenta una revisión detallada de la situación actual de las evaluaciones docentes en 19 países de la región de América Latina y el Caribe. La conclusión que se desprende de este estudio es que existen cinco claves para una evaluación docente exitosa.

Primero, el diseño y la implementación de cualquier tipo de política docente, particularmente aquellas relacionadas con la evaluación, requieren claridad con respecto a lo que

se espera de un docente de excelencia, y exigen que quede definido lo que el docente debe *saber* y debe *saber hacer*. Es decir, es preciso contar con **estándares de desempeño docente** que determinen de manera detallada las características y habilidades que un docente debe tener tanto en términos pedagógicos como de conocimientos disciplinares. Estos estándares son útiles además para los formuladores de políticas, ya que les permiten regular otros ejes de la profesión —incluidos los de formación inicial y en servicio—, e incluso para el propio docente, de forma que él mismo conozca cuales son las expectativas que existen en cuanto a sus conocimientos y capacidades. El análisis permite concluir que tan solo cinco de los 19 países analizados han desarrollado de manera detallada este tipo de estándares, a través de procesos consensuados con distintos actores del sector educativo. Es importante que los demás países de la región avancen en esta dirección fundamental para el sector educativo y definan para sus propios contextos el docente ideal que requieren sus sistemas.

Segundo, es indispensable tener en claro **qué objetivo se desea alcanzar con la implementación de las evaluaciones docentes**. Las evaluaciones docentes pueden tener dos fines generales: i) mejorar las prácticas y habilidades pedagógicas y/o disciplinares de los docentes, y ii) mejorar la composición y motivación de la fuerza docente a través de incentivos positivos o negativos. Lograr ambos objetivos con una única herramienta o tipo de evaluación es difícil, ya que los docentes estarán más o menos dispuestos a revelar sus verdaderas falencias o habilidades dependiendo del objetivo buscado. Una evaluación que tiene un objetivo formativo puede permitir que el docente actúe con mayor naturalidad y honestidad a la hora de ser evaluado, porque sabe que esto puede traducirse en un apoyo para mejorar su práctica docente. Por otro lado, una evaluación que tiene consecuencias laborales puede llevar a que el docente enfoque su práctica de enseñanza exclusivamente en los aspectos a ser evaluados para poder lograr los beneficios o evitar las consecuencias negativas.

Por lo tanto, no es de extrañar que ciertos casos analizados —tanto en los sistemas educativos de alto desempeño como en algunos de la región— recurran a la implementación de diferentes tipos de evaluaciones, en función del tramo de la carrera en el que se encuentre el docente y del objetivo que buscan alcanzar. El análisis sugiere que esta división de objetivos se debe fortalecer en los países de la región en donde ya existe, e introducir en aquellos en donde aún no se han puesto en marcha evaluaciones que diferencien claramente esta dupla de objetivos. Más aún: sería deseable que las evaluaciones docentes obligatorias y periódicas se encaminaran a detectar las principales debilidades pedagógicas y disciplinares de los maestros, y a superarlas a través de la capacitación y formación continua. De manera complementaria, se deberían llevar adelante evaluaciones cuyos resultados se aten a ascensos, bonificaciones o incluso remoción de las aulas en caso de ser necesario. Estas se podrían realizar con menor periodicidad y con la utilización de otros instrumentos complementarios que evalúen aspectos adicionales del desempeño docente.

Tercero, sin importar el tipo de evaluación que los formuladores de políticas busquen implementar —de ingreso, de período de prueba, periódica obligatoria o de ascenso o bonificaciones—, para que las evaluaciones en realidad capturen la efectividad de los docentes, se debe asegurar que su **diseño e implementación otorgue resultados confiables y válidos**. Es decir, que deben reflejar de manera consistente aspectos exclusivamente relacionados con las prácticas docentes y no estar sujetas a otros factores (por ejemplo, no deben depender de quién es el evaluador o del día en que se hace la evaluación), y que deben estar altamente correlacionadas con los aprendizajes de los estudiantes y el desempeño estudiantil. Estas características son indispensables para que la evaluación docente pueda otorgar información útil y relevante a los docentes y al sistema educativo en general, y para que constituya una medida aceptada por los docentes como un buen indicador de su desempeño. La complejidad de la labor docente implica a su vez que una evaluación válida y confiable cuente con el uso de una multiplicidad de instrumentos que permitan capturar algunas de las numerosas tareas que ejercen los docentes en las aulas. Estas tareas incluyen, entre otras cosas, la promoción efectiva de las habilidades cognitivas y socioemocionales de los estudiantes (véanse, por ejemplo, Jackson, 2018 y Kraft, 2019).

La evidencia de uno de los proyectos más completos en el área (MET, 2013) sugiere que una combinación de tres instrumentos —observaciones estandarizadas en el aula, medidas de valor agregado en base a exámenes estandarizados de los estudiantes, y encuestas a los propios estudiantes— puede resultar efectiva para detectar maestros sobresalientes en la promoción de los aprendizajes de los estudiantes en Estados Unidos. El uso de estos instrumentos, así como el peso que puedan tener en la evaluación dependerán, entre otras cosas, del objetivo principal que esta última persiga, de los estándares de enseñanza que existan en cada país, del estado de la política docente, de la calidad de la formación inicial y en servicio, y del presupuesto disponible para llevar a cabo las evaluaciones. Además, como lo han demostrado algunos estudios sobre el tema (Sartain y Steinberg, 2015; Stecher et al., 2018), la calidad de la implementación de los instrumentos finalmente escogidos es también fundamental y determinará la posibilidad de que en efecto se alcancen los objetivos esperados de los sistemas de evaluación.

Para lograr que la evaluación docente otorgue resultados válidos y confiables, la elección de los instrumentos usados para medir la efectividad es esencial. Al respecto, vale la pena destacar algunos aprendizajes relacionados con el uso y el diseño de cuatro de los instrumentos más comúnmente considerados para medir la efectividad de los docentes en promover el desarrollo de las habilidades de sus estudiantes: medidas de valor agregado docente, rúbricas de observación en clase, portafolios docentes y exámenes de conocimiento disciplinar.

Con respecto a las **medidas de valor agregado docente** —las cuales se basan en los resultados de múltiples evaluaciones estandarizadas de estudiantes, generalmente administradas al principio y al final de un grado o año escolar—, la evidencia sugiere que, aunque son adecuadas para identificar quiénes son los maestros más (o menos) efectivos, resultan

difíciles y costosas de implementar, ya que se deben diseñar y administrar pruebas de conocimiento a todos los estudiantes como mínimo al finalizar cada año escolar, o incluso al comienzo de este también. El nivel de inversión que su implementación demanda es tan elevado que incluso en países que disponen de mayores recursos, como es el caso de Estados Unidos, tan solo entre el 20%-25% de los docentes cuenta con información suficiente para construir esta medida. Asimismo, estas medidas no necesariamente ofrecen la información que se requiere para mejorar la efectividad de los docentes, ya que no detallan muchos aspectos de lo que el docente está realizando o no en clase. Si bien se trata de instrumentos que permiten identificar la efectividad de los docentes en promover el aprendizaje entre sus estudiantes, no necesariamente proveen información sobre qué es lo que el docente está haciendo bien o mal, o de manera insuficiente dentro del aula de clase, lo que limita la posibilidad de identificar las fortalezas y debilidades de los maestros, y ofrecer acciones encaminadas a la mejora de su desempeño.

Existen otros instrumentos adicionales que pueden ser muy útiles para identificar las fortalezas y debilidades de los docentes, y que pueden servir como complemento de las medidas de valor agregado. Entre ellos se encuentran las **herramientas de observación en clase**. El análisis sugiere que fortalecer los esfuerzos para implementar y mejorar el uso de este instrumento sería beneficioso para la región. Las deficiencias en las prácticas pedagógicas de los docentes de América Latina y el Caribe se evidencian en los resultados de los sistemas de evaluación en países como Chile, Ecuador o República Dominicana, así como también en estudios regionales que han efectuado observaciones de la práctica docente en varios países de la región (Bruns y Luque, 2015).

Las rúbricas o los formularios usados para la observación en el aula consideran desde el uso del tiempo de clase por parte del docente (e.g., Stallings) hasta la calidad de las interacciones del docente con sus estudiantes (e.g., CLASS). Pero pueden medir diferentes aspectos de lo que sabe y sabe hacer el maestro. A pesar de que la observación en clase es el instrumento más utilizado en los sistemas que se analizan en este estudio, su implementación es compleja. Como se mencionó anteriormente, la dificultad para usar este tipo de instrumentos radica en generar observaciones en el aula que sean sostenibles y escalables, debido tanto al costo de la implementación de estas rúbricas, como la capacidad instalada necesaria para llevarlas a cabo. Además, varios elementos de su diseño pueden influir en su efectividad: es necesario definir cuántas observaciones se deben efectuar, si se deben aplicar de manera presencial o a través de videos, quién será el encargado de llevarlas a cabo y calificarlas, si deben ser previamente anunciadas o ejecutarse por sorpresa, entre otros.

Todas estas decisiones no tienen una respuesta única, y el diseño y la implementación del instrumento dependerá del contexto en el que se vaya a llevar a cabo, y de los recursos disponibles para ello. Por ejemplo, si se decide realizar la observación de clases in situ, dada la baja capacidad instalada que existe en la región para aplicar este tipo de herramientas, junto con las falencias antes mencionadas en la práctica docente, habría que comenzar por emplear rúbricas sencillas, escalables y sostenibles (Bruns, De Gregorio

y Taut, 2016). Estas permitirán identificar las falencias más básicas en el desempeño del docente y aquellas urgentes de superar, y posibilitarán en el tiempo la utilización de rúbricas más completas. Si, por el contrario, lo que se decide es la observación de prácticas a través de videos, sería relativamente más sencillo implementar rúbricas más completas, que midan aspectos adicionales de la labor docente. Sin embargo, es fundamental definir cómo se realizarían los videos, quién lo haría, y si serían responsabilidad del propio docente o de personas externas a la escuela.

El uso de **portafolios docentes** que incentivan al maestro para que lleve a cabo procesos de autorreflexión y permiten la recolección de evidencia directa en cuanto a cómo planifican, aplican y evalúan los conocimientos en su área, ha demostrado ser valioso tanto en el contexto latinoamericano como en países desarrollados. De hecho, esta herramienta ya está siendo utilizada en Chile y Honduras, países de cuya experiencia se pueden desprender aprendizajes importantes en cuanto a su estructura e implementación. Además, se trata de una herramienta alternativa que permite identificar algunas de las habilidades pedagógicas en sistemas con un gran número de docentes, con características geográficas difíciles, o con restricciones presupuestarias, que hacen que sea complicado llevar adelante una evaluación censal usando un instrumento de observación en el aula. Por supuesto, como todos los instrumentos de evaluación, su implementación es fundamental. Por ejemplo, la calificación de dichos portafolios debe quedar en manos de personal capacitado, y para que esta calificación no se produzca de manera subjetiva, debe contar con criterios objetivos, claros y sistemáticos. De la misma forma, se debe garantizar que no exista la posibilidad de creación de “mercados” para la elaboración de estos portafolios por alguien distinto al mismo maestro, y que sean verdaderamente el resultado del trabajo de reflexión de cada docente.

Las evaluaciones docentes también podrían incluir **exámenes disciplinares** que permitan determinar cuál es el nivel de conocimientos que cada maestro tiene en su área específica. Al centrarse en las habilidades y saberes pedagógicos, algunos de los instrumentos arriba referenciados no permiten identificar falencias en este aspecto fundamental de la docencia. Evaluar el conocimiento disciplinar de los docentes es relativamente menos complicado y costoso que evaluar habilidades pedagógicas. Pero, como toda política docente, su diseño e implementación no son triviales, y se requiere el desarrollo de exámenes de conocimientos que concuerden con los objetivos de aprendizaje de cada área y grado escolar en el sistema.

Estos exámenes disciplinares deberían incluirse como parte de los diferentes tipos de evaluación docente. Primero, como parte de los exámenes de ingreso a la carrera docente, para que aseguren que solo puedan ingresar y ejercer la profesión aquellos individuos con un cierto nivel de conocimientos. La pérdida del prestigio de la profesión, documentada por Elacqua et al. (2018), se explica en parte por la disminución de la calidad de la formación inicial, lo que a su vez trae implicaciones en cuanto al nivel de conocimiento disciplinar de los docentes. Por ello, también es necesario mejorar la calidad de los programas de

formación inicial y continuar atrayendo a candidatos con mayores niveles de desempeño para que estudien e ingresen a la carrera docente. Un segundo momento en el cual se podrían implementar estas pruebas es al tomar exámenes de ascenso, tal y como se hace en Chile o Perú, para asegurar que los profesionales con mayores niveles en el escalafón docente tengan el dominio disciplinar y pedagógico que ese nivel del escalafón requiere. Esto aseguraría que, en caso de que la reglamentación así lo permita, los docentes con altos niveles de conocimiento y, por supuesto, con prácticas pedagógicas de excelencia, puedan ejercer como tutores o mentores de aquellos que así lo requieran.

Otro instrumento que, en algunos casos, ha resultado útil para identificar la efectividad de los docentes (por ejemplo, en el proyecto de MET), ha sido la **encuesta a los estudiantes** sobre aspectos relacionados a la calidad docente. Sin embargo, la aplicación de dicho instrumento es muy sensible respecto al tipo y forma de las preguntas, ya que estas pueden resultar difíciles de entender particularmente para los estudiantes más jóvenes. En otras ocasiones, las encuestas a los estudiantes han sido acompañadas por encuestas a padres, sin embargo, no hay evidencia sobre su efectividad para identificar la calidad de los docentes.

La cuarta clave para una evaluación docente exitosa es que **los resultados que surjan de estas evaluaciones deben ser utilizados en la toma de decisiones laborales y de formación continua de los maestros**. Es indispensable hacer un uso efectivo y adecuado de la información proveniente de los resultados, ya que solo a través de verdaderas transformaciones en la capacitación continua y en la composición de la fuerza docente será posible mejorar la calidad de la enseñanza y los aprendizajes de manera sistémica. Si el objetivo principal del sistema de evaluación es la mejora del conocimiento pedagógico y/o disciplinar del docente, la evaluación debe venir acompañada de programas de formación en servicio que respondan de manera directa a las debilidades que las evaluaciones reflejen. En la región, tan solo las regulaciones de Chile y Perú atan los resultados de las evaluaciones obligatorias periódicas a este tipo de programas. Aunque se han llevado a cabo algunas iniciativas destacadas aún falta mucho camino por recorrer en este aspecto en la región (Cortés et al., 2011; Schmelkes, 2017). La planificación y puesta en marcha de este tipo de programas también son complejas y costosas, y demandan inversiones elevadas en tiempo y dinero, factores que podrían explicar los escasos avances que ha tenido la región en este frente. Sin embargo, queda claro que solo a través de este tipo de programas será posible mejorar las capacidades de la fuerza docente en los países de América Latina y el Caribe.

El uso de la información para otorgar incentivos laborales tampoco es sencillo y el diseño de un sistema de bonificaciones, ascensos o despidos dentro de la carrera debe ser cuidadoso para no llevar a incentivos perversos. Por ejemplo, al aportar información de aquellos maestros más o menos efectivos, las medidas de valor agregado pueden ser utilizadas para generar cambios en la composición de la fuerza docente a través de la remoción

de aquellos que de manera consistente sean poco efectivos o no estén promoviendo el aprendizaje de sus estudiantes. Lo mismo puede aplicarse a los resultados obtenidos a través de herramientas de observación en clase, portafolios o exámenes disciplinares que sugieran que el docente no tiene las habilidades pedagógicas ni el conocimiento suficiente que tan importante labor requiere. Todos estos resultados pueden estar atados también a ascensos, bonificaciones u otro tipo de incentivos. Sin embargo, la evidencia sugiere que lo fundamental es que el diseño de estos incentivos sea muy cuidadoso para asegurar que estos logren mejoras significativas en los aprendizajes estudiantiles, y que no se conviertan simplemente en estrategias de enseñanza solo para los exámenes (“*teach to the test*”) o que puedan incentivar la copia o trampa por parte de los docentes para aumentar los puntajes y obtener así los beneficios que de ellos puedan surgir.

Al definir el uso que se les dará a los resultados de las evaluaciones, se debe tener en cuenta también la economía política asociada a la implementación de estas medidas. Por ejemplo, en función del uso de los resultados finalmente definidos por las autoridades podría generarse un rechazo por parte de la fuerza docente si los docentes perciben al sistema de evaluación más como una amenaza que como una alternativa de mejoramiento continuo en su profesión. La aceptación o no del diseño final de la evaluación y de sus consecuencias por parte de los distintos actores dentro del sistema educativo podría fortalecer o minar los esfuerzos de los formuladores de políticas (Elacqua et al., 2018; Bruns et al., 2019). Aunque se trata de un tema que excede el alcance de este estudio, vale la pena resaltar que una de las experiencias más ricas en términos de este manejo político de la evaluación docente proviene de Chile, país donde las implicaciones laborales atadas a la evaluación fueron agregadas de manera gradual después de mucho tiempo y de extensas negociaciones con el sindicato, cuando los maestros ya entendían y valoraban la evaluación, y parte importante de dicho sistema ya estaba en funcionamiento (Mizala y Schneider, 2014).

La quinta clave para la implementación de un sistema de evaluación exitoso, que otorgue información válida y confiable, es la **investigación y revisión continua de los instrumentos, procesos, resultados e impactos** que de ella surjan. La decisión final de los instrumentos incluidos en una evaluación, al igual que su implementación en el campo, no es trivial. Además, la puesta en marcha de sistemas de evaluación exitosos requiere un gran esfuerzo en términos de tiempo y dinero por parte de todos los actores que se hallan dentro del sistema educativo. Por lo tanto, la investigación es importante para asegurar que haya validez en el proceso de elaboración de los instrumentos (que estos estén alineados efectivamente con los estándares de desempeño), que las evaluaciones provean información válida y confiable, y que se trate de un sistema a favor de la mejora de la calidad educativa. Aunque los aprendizajes existentes con respecto a los distintos instrumentos a ser incluidos son valiosos, estos provienen en su gran mayoría de sistemas educativos muy distintos de los que existen en la región, por lo que el uso y la adaptabilidad a este contexto es aún una tarea

pendiente. Además, como se detalló en el capítulo 4, la investigación alrededor de los instrumentos, procesos y requerimientos de los sistemas de evaluación de países como Chile ha permitido que estos hayan ido cambiando y mejorando en el tiempo. Por tanto, estos esfuerzos de investigación permanente no solo influirían de manera positiva en el correcto funcionamiento de la evaluación, y en la aceptación de los diversos actores del sistema, sino que también contribuirían a complementar la escasa literatura que hay acerca del tema a nivel regional e internacional.

Los países de América Latina aún tienen un largo camino que recorrer para implementar evaluaciones docentes sistémicas que sean confiables, válidas, y cuyos resultados se usen de manera efectiva para mejorar la calidad de los docentes. Varios países han dado pasos importantes, con evaluaciones que incluyen la observación de la práctica docente en el aula, pruebas de conocimiento o disciplinares, o el uso de portafolios. En cada caso se han usado diferentes instrumentos, que se han implementado de formas variadas y con resultados diversos. **Aun así, todos coinciden en algo: solo una evaluación docente exitosa logrará mejorar la enseñanza y los aprendizajes de manera sistémica y abrir el camino hacia una mejor calidad educativa para todos los niños, niñas y jóvenes de nuestra región.**

REFERENCIAS BIBLIOGRÁFICAS

- Aaronson, D., L. Barrow y W. Sander. 2007. Teachers and Students Achievement in the Chicago Public High Schools. *Journal of Labor Economics*, 25(1): 95-135.
- Adnot, M., T. Dee, V. Katz y J. Wyckoff. 2017. Teacher Turnover, Teacher Quality, and Student Achievement in DCPS. *Educational Evaluation and Policy Analysis*, 39(1): 54-76.
- Alvarado, M., F. Duarte, C. Neilson y Mineduc (Ministerio de Educación de Chile). 2012. Efectos preliminares de la Beca Vocación de Profesor. Santiago de Chile: Centro de Estudios, Ministerio de Educación de Chile (Mineduc).
- American Educational Research Association, American Psychological Association, y National Council on Measurement in Education. 2014. Standards for Educational and Psychological Testing. Washington, D.C.: American Psychological Association (APA).
- Aparecido Cordão, F. 2012. Carreira docente: análise jurídica e institucional. Trabajo encargado por el BID. (Documento sin publicar).
- Araujo, M. C., P. Carneiro, Y. Cruz-Aguayo y N. Schady. 2016. Teacher Quality and Learning Outcomes in Kindergarten. *The Quarterly Journal of Economics*, 131 (3): 1415-53.
- Ávalos, B. 2016. "Learning from Research on Beginning Teachers". En: J. Loughran y M. L. Hamilton (eds.), *International Handbook of Teacher Education*. Nueva York: Springer.
- Banco Mundial. 2017. *Indicadores del Desarrollo Mundial*. Washington, D.C.: Banco Mundial.
- . 2018. *World Development Report 2018: Learning to Realize Education's Promise*. Washington, D.C.: Banco Mundial.
- Banerjee, Abhijit Vinayak, Shawn Cole, Esther Duflo, and Leigh Linden. 2007. Remedying Education: Evidence from Two Randomized Experiments in India. *Quarterly Journal of Economics* 122 (3): 1235-64.

- Banuri, S., P. Keefer y D. de Walque. 2018. Love the Job... or the Patient? Task vs. Mission-Based Motivations in Health Care. Documento de trabajo Núm. 8338. Washington, D.C.: Banco Mundial.
- Barrera-Osorio, F. y D. Raju. 2017. Teacher performance pay: Experimental evidence from Pakistan. *Journal of Public Economics, Elsevier*, Vol. 148(C): 75-91.
- Benton, S. L. y W. E. Cashin. 2012. Student ratings of teaching: A summary of research and literature. Documento de ideas Núm. 50. IDEA Centre.
- Bertoni, E., G. Elacqua, A. Jaimovich, J. Rodríguez y H. Santos. 2018. Teacher Policies, Incentives, and Labor Markets in Chile, Colombia, and Peru: Implications for Equality. Documento de trabajo Núm. 9124. Washington, D.C.: BID.
- BID (Banco Interamericano de Desarrollo)-CIMA (Centro de Información para la Mejora de los Aprendizajes). 2019. Tasa de asistencia por grupo etario. Disponible en: <https://www.iadb.org/es/cima>.
- Bruns, B. y J. Luque. 2015. Great Teachers: How to Raise Student Learning in Latin America and the Caribbean. Washington, D.C.: Banco Mundial.
- Bruns, B., S. De Gregorio y S. Taut. 2016. Measures of Effective Teaching in Developing Countries. RISE-WP-16/009.
- Bruns, B., I. H. Macdonald y B.R. Schneider. 2019. The politics of quality reforms and the challenges for SDGs in education. *World Development*, Elsevier, Vol. 118(C): 27-38.
- Brutti, Z. y F. Sánchez. 2017. Does Better Teacher Selection Lead to Better Students? Evidence from a Large-Scale Reform in Colombia. Documento CEDE Núm. 015350. Bogotá: Universidad de los Andes.
- Busso, M., J. Cristiá, D. Hincapié, J. Messina y L. Ripani. 2017. *Aprender mejor: políticas públicas para el desarrollo de habilidades*. Serie Desarrollo en las Américas. Washington, D.C.: BID.
- Campbell, D. 1979. Assessing the impact of planned social change. *Evaluation and Program Planning*, 2(1): 67-90.
- Caraballo, D., A. García, M. E. Báez y F. Vargas. 2014. Estado de Situación de las Políticas Docentes en República Dominicana. Documento de trabajo PREAL/EDUCA.
- Cevallos-Estarellas, P. 2016. Avances en las políticas docentes en Ecuador entre 2007 y 2016. Trabajo de consultoría realizado para la oficina del Banco Interamericano de Desarrollo, Ecuador.

- Chetty, R., J. N. Friedman y J. E. Rockoff. 2014a. Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates. *American Economic Review*, 104(9), 2593-2632. Disponible en: <https://doi.org/10.1257/aer.104.9.2593>.
- . 2014b. Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood. *American Economic Review*, 104(9), 2633-79. Disponible en: <https://doi.org/10.1257/aer.104.9.2633>.
- Chingos, M. M. y P. E. Peterson. 2011. It's Easier to Pick a Good Teacher than to Train One: Familiar and New Results on the Correlates of Teacher Effectiveness. *Economics of Education Review*, 30 (3): 449-65.
- Cohen, J. y D. Goldhaber. 2016. Building a More Complete Understanding of Teacher Evaluation Using Classroom Observations. *Educational Researcher*, 45(6): 378-387.
- Cortés, F., S. Taut, M. V Santelices y M. J. Lagos 2011. Formación continua en profesores y la experiencia de los Planes de Superación Profesional (PSP) en Chile: Fortalezas y debilidades a la luz de la evidencia internacional. Segunda Reunión Anual de la Sociedad Chilena de Políticas Públicas, Santiago, Chile.
- CUREE (Centre for the Use of Research and Evidence in Education). 2018. International Review of Teacher Evaluation Systems: Technical Report and Comparative Analysis. Coventry: CUREE. Documento de antecedentes preparado para el proyecto del BID "Measuring teacher effectiveness in Latin America", 2017-18.
- Dee, T. S. y J. Wyckoff. 2015. Incentives, Selection, and Teacher Performance: Evidence from IMPACT. *Journal of Policy Analysis and Management*, John Wiley & Sons, Ltd., Vol. 34(2), pp. 267-297 (marzo).
- Docentemás. 2017. Portafolio. Uso Formativo Santiago de Chile: Ministerio de Educación. Disponible en: <http://www.docentemas.cl/pages/portafolio/uso-formativo.php>.
- Donaldson, M. L. y J. P. Papay. 2015. "Teacher evaluation for accountability and development." En: H. F. Ladd y M. E. Goertz (eds.), *Handbook of Research in Education Finance and Policy*, pp. 174-193. Nueva York: Routledge.
- Elacqua, G., D. Hincapié, E. Vegas y M. Alfonso. 2018. Profesión: Profesor en América Latina. ¿Por qué se perdió el prestigio docente y cómo recuperarlo? Washington, D.C.: BID.
- Estrada, 2019. Rules versus Discretion in Public Service: Teacher Hiring in Mexico. *Journal of Labor Economics*, 37(2), 545-579.

- Evans, D. K. y A. Popova. 2016. What Really Works to Improve Learning in Developing Countries? An Analysis of Divergent Findings in Systematic Reviews. Oxford University Press. Washington, D.C.: Banco Mundial. Disponible en: <https://openknowledge.worldbank.org/handle/10986/29308>.
- FEREMA (2015). Honduras el Estado de las Políticas Públicas Docentes. Tegucigalpa: Informes de Seguimiento PREAL.
- Figuerola, M., S. García, D. Maldonado, C. Rodríguez, A.M. Saavedra, G. Vargas. 2018. La profesión docente en Colombia: normatividad, formación, selección y evaluación. Documento de trabajo Núm. 54. Bogotá: Escuela de Gobierno, Universidad de los Andes.
- Fryer, R. 2013. Teacher Incentives and Students Achievement: Evidence from New York City Public Schools. *Journal of Labor Economics*, 31(2): 373–427.
- Ganimian, A. J. y R. J. Murnane. 2016. Improving education in developing countries: Lessons from rigorous impact evaluations. *Review of Educational Research*, 86(3), 719–755. Disponible en: <http://dx.doi.org/10.3102/0034654315627499>.
- García, S., D. Maldonado, G. Perry, C. Rodríguez y J. E. Saavedra. 2014. *Tras la excelencia docente: Como mejorar la calidad educativa de todos los colombianos*. Buenos Aires: Fundación Compartir.
- Gertler, P. J., S. Martínez, P. Premand, L. B. Rawlings y C. M. J. Vermeersch. 2016. *Impact Evaluation in Practice (segunda edición)*. Washington, D.C.: BID y Banco Mundial. Disponible en: <https://openknowledge.worldbank.org/handle/10986/25030>.
- Goldhaber, D. y E. Anthony. 2007. Can teacher quality be effectively assessed? National board certification as a signal of effective teaching. *The Review of Economics and Statistics*, 89(1): 134–150.
- Hanushek, E. A., J. F. Kain, D. M. O'Brien y S. G. Rivkin. 2005. The Market for Teacher Quality. Documento de trabajo del NBER Núm. 11154. Cambridge, MA: National Bureau of Economic Research.
- Hanushek, E. A. y S. G. Rivkin. 2010. Generalizations about Using Value-Added Measures of Teacher Quality. *American Economic Review*, 100 (2): 267–71.
- Harris, D. N. 2011. Value-Added Measures in Education. Cambridge, MA: Harvard University Press.
- INEE (Instituto Nacional para la Evaluación de la Educación). 2018. La educación obligatoria en México, Informe 2018. Ciudad de México: INEE.

- INEEd (Instituto Nacional de Evaluación Educativa de Uruguay). 2017. Informe sobre el estado de la educación en Uruguay 2015-2016. Montevideo: INEEEd.
- Ingersoll, R. M. y M. Strong. 2011. The Impact of Induction and Mentoring Programs for Beginning Teachers: A Critical Review of the Research. *Review of Educational Research*, 81 (2): 201-233.
- Jackson, C. K. 2018. What Do Test Scores Miss? The Importance of Teacher Effects on Non-Test Score Outcomes. *Journal of Political Economy*, 126, Núm. 5 (octubre): 2072-2107.
- Jensen, B., A. Sandoval-Hernández, S. Knoll y E. González. 2012. The Experience of New Teachers: Results from TALIS 2008. París: OCDE.
- Kane, T., K. Kerr y R. Pianta. 2014. *Designing teacher evaluation systems: New guidance from the measures of effective teaching project*. Hoboken, NJ: John Wiley & Sons.
- Kane, T. J., D. F. McCaffrey, T. Miller, T. y D. O. Staiger. 2013. Have We Identified Effective Teachers? Validating Measures of Effective Teaching Using Random Assignment. Documento de investigación. Seattle: MET Project, Bill & Melinda Gates Foundation. Disponible en: <https://eric.ed.gov/?id=ED540959>.
- Koretz, D. 2015. Adapting Educational Measurement to the Demands of Test-Based Accountability. *Measurement: Interdisciplinary Research and Perspectives*, 13(1): 1-25.
- Kraft, M. A. 2019. Teacher effects on complex cognitive skills and social-emotional competencies. *Journal of Human Resources*, 54(1): 1-36.
- Kraft, M. A. y A. F. Gilmour. 2017. Revisiting the Widget Effect: Teacher Evaluation Reforms and the Distribution of Teacher Effectiveness. *Educational Researcher*, 46(5): 234-249.
- Kremer, M., C. Brannen y R. Glennerster. 2013. The challenge of education and learning in the developing world. *Science*, 340(6130): 297-300.
- Lachlan-Haché, L., L. Matlach, A. Guiden y M. Castro. 2015. What We Know About SLOs: An Annotated Bibliography of Research on and Evaluations of Student Learning Objectives. Washington, D.C.: American Institutes for Research.
- Lai, F., L. Zhang, Q. Qinghe, X. Hu, Y. Shi, M. Boswell y S. Rozelle. 2012. Does Computer-Assisted Learning Improve Learning Outcomes? Evidence from a Randomized Experiment in Public Schools in Rural Minority Areas in Qinghai, China. Documento de trabajo REAP Núm. 237, Rural Education Action Program, Freeman Spogli Institute. Stanford, CA: Stanford University.

- Lockwood, J., D. F. McCaffrey, L. S. Hamilton, B. Stecher, V. N. Le y J. F. Martínez. 2007. The sensitivity of value-added teacher effect estimates to different mathematics achievement measures. *Journal of Educational Measurement*, 44(1): 47-67.
- Luschei, T. F., A. Chudgar y W. J. Rew. 2013. Exploring Differences in the Distribution of Teacher Qualifications across Mexico and South Korea: Evidence from the Teaching and Learning International Survey Teachers College Record, v115 n5 2013.
- Marcelo, C. y D. Vaillant. 2017. Políticas y Programas de Inducción en la Docencia en Latinoamérica. *Cuadernos de Pesquisa*, Vol. 74, pp. 1224-29.
- McEwan, P. J. 2015. Improving Learning in Primary Schools of Developing Countries: A Meta-Analysis of Randomized Experiments. *Review of Educational Research*, 85(3): 353-394. Disponible en <https://doi.org/10.3102/0034654314553127>.
- MEC (Ministerio de Educación y Cultura de Paraguay). 2009. Reglamento del Concurso Público de Oposición para la Selección de Educadores. Aprobado por Resolución Ministerial N° 9135 del 12 de octubre de 2009. Asunción: MEC.
- . 2016. Resolución 26704 del 1 de diciembre de 2016. Asunción: MEC.
- MET. 2013. Ensuring Fair and Reliable Measures of Effective Teaching: Culminating Findings from the MET Project's Three-Year Study. Policy and Practice Brief. Seattle: MET Project, Bill & Melinda Gates Foundation.
- Mizala, A. y B. Ross Schneider. 2014. Negotiating Education Reform: Teacher Evaluations and Incentives in Chile (1990-2010). *Governance: An International Journal of Policy, Administration, and Institutions*, Vol. 27(1): 87-109.
- . 2019. Promoting quality education in Chile: the politics of reforming teacher careers. *Journal of Education Policy*, DOI: 10.1080/02680939.2019.1585577.
- Molina, E. et al. 2018. Measuring Teaching Practices at Scale: Results from the Development and Validation of the Teach Classroom Observation Tool. Documento de trabajo de investigación de políticas Núm. 8653. Washington, D.C.: Banco Mundial.
- Muralidharan, K., A. Singh y A. Ganimian. 2019. Disrupting education? Experimental evidence on technology-aided instruction in India. *American Economic Review*, 109(4): 1426-1460. Disponible en <https://doi.org/10.1257/aer.20171112>
- Murnane, R. J. y A. J. Ganimian. 2014. Improving educational outcomes in developing countries: Lessons from rigorous evaluations. Documento de trabajo del NBER Núm. 20284. Cambridge, MA: National Bureau of Economic Research.

- Näslund, E., H. Alonzo y D. Martin. 2013. Challenges and Opportunities in the Belize Education Sector. Nota técnica Núm. IDB-TN-538. Washington, D.C.: BID.
- Nogueira, D. X. P., G. Ribeiro de Jesus y S. Pereira da Silva Cruz. 2013. Avaliação de desempenho docente no Brasil: desvelando concepções e tendencias. *Linhas Críticas*, Brasília, DF, v. 19, n. 38, pp. 13-32.
- OCDE (Organización para la Cooperación y el Desarrollo Económicos). 2009. Teacher Evaluation: A Conceptual Framework and Examples of Country Practices. París: OCDE.
- . 2013. Teachers for the 21st Century: Using Evaluation to Improve Teaching. París: OCDE.
- . 2015. Who Wants to Become a Teacher? PISA in Focus 58. París: OCDE.
- . 2016. Reviews of School Resources: Uruguay 2016. OECD Reviews of School Resources. París: OCDE.
- . 2017. Education in Costa Rica, Reviews of National Policies for Education. París: OCDE.
- Ome, A. 2012. The Effects of Meritocracy for Teachers in Colombia. Informe de investigación Núm. 010260. Bogotá: Fedesarrollo.
- Papay, J. P. 2011. Different Tests, Different Answers the Stability of Teacher Value-Added Estimates across Outcome Measures. *American Educational Research Journal*, 48: 163-193.
- Papay, J., Taylor, E. S., Tyler, J. y M. Laski. 2016. Learning Job Skills from Colleagues at Work: Evidence from a Field Experiment Using Teacher Performance Data. Documento de trabajo del NBER Núm. 21986. Cambridge, MA: National Bureau of Economic Research.
- Polikoff, M. S. 2015. The Stability of Observational and Student Survey Measures of Teaching Effectiveness. *American Journal of Education*, 121(2): 183-212.
- Rivkin, S. G., E. A. Hanushek y J. F. Kain. 2005. Teachers, Schools, and Academic Achievement. *Econometrica*, 73(2): 417-458. Disponible en: <https://doi.org/10.1111/j.1468-0262.2005.00584.x>.
- Rockoff, J. E. 2004. The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data. *The American Economic Review*, 94(2): 247-252.
- Rockoff, J. E. y C. Speroni. 2010. Subjective and Objective Evaluations of Teacher Effectiveness. *The American Economic Review: papers and Proceedings*, Vol. 100, Núm. 2.

- Rockoff, J. E., D. O. Staiger, T. J. Kane y E. S. Taylor. 2012. Information and Employee Evaluation: Evidence from a Randomized Intervention in Public Schools. *American Economic Review* 2012, 102(7): 3184–3213.
- Rodríguez, C. 2015. El manejo docente en la República Dominicana: análisis y recomendaciones de política a partir de un estudio comparativo con países exitosos y algunos de América Latina. Documento de antecedentes preparado para el proyecto del BID “Measuring teacher effectiveness in Latin America”, 2017–18. Washington, D.C.: BID.
- Rodríguez, C. y M. C. Rivera. 2018. Sistemas de evaluación docente en diecinueve países de América Latina y el Caribe. Documento de consultoría. Washington, D.C.: BID.
- Santibáñez, L., J. F. Martínez, A. Datar, P. J. McEwan, C. Messan-Setodji y R. Basurto-Dávila. 2006. Haciendo camino: Análisis del sistema de evaluación y del impacto del programa de estímulos docentes Carrera Magisterial en México. Ciudad de México: Secretaría de Educación Pública.
- Santibáñez, L., D. Rubio y M. Vázquez. 2017. Formación Continua de Docentes: Política Actual en México y Buenas Prácticas Nacionales e Internacionales. Reporte final presentado al Banco Interamericano de Desarrollo.
- Sartain, L. y M. P. Steinberg. 2016. Teachers’ Labor Market Responses to Performance Evaluation Reform: Experimental Evidence from Chicago Public Schools. *Journal of Human Resources*, Vol. 51(3): 615–655.
- Schleicher, A. 2016. Teaching excellence through professional learning and policy reform. Lessons from Around the World, International Summit on the Teaching Profession.
- Schmelkes, S. 2017. “La evaluación docente en México”. Presentación en el Instituto Nacional Para la Evaluación de la Educación, México.
- Smith, W. C. y K. Kubacka. 2017. The emphasis of student test scores in teacher appraisal systems. *Education Policy Analysis Archives*, 25.
- Springer, M. G. (ed.). 2009. *Performance Incentives*. Washington, D.C.: Brookings Institution Press.
- Stallings, J. y S. Knight. 2003. Using the Stallings Observation System to Investigate Time on Task in Four Countries. Documento no publicado preparado para el proyecto International Time on Task (ITOT). Washington, D.C.: Banco Mundial.

- Stecher, B. 2002. Consequences of large-scale, high-stakes testing on school and classroom practice. En: L. S. Hamilton, B. M. Stecher y S. P. Klein (eds.), *Making sense of test-based accountability in education*, pp. 79-100. Santa Mónica, CA: RAND.
- Stecher, B. M., D. J. Holtzman, M. S. Garet, L. S. Hamilton, J. Engberg, E. D. Steiner, A. Robyn, M. D. Baird, I. A. Gutiérrez, E. D. Peet, I. Brodziak de los Reyes, K. Fronberg, G. Weinberger, G. P. Hunter y J. Chambers. 2018. Improving Teaching Effectiveness: Final Report: The Intensive Partnerships for Effective Teaching Through 2015-2016. Santa Monica, CA: RAND Corporation.
- Steele, J., M. Pepper, M. Springer y J. R. Lockwood. 2015. The distribution and mobility of effective teachers: Evidence from a large, urban school district. *Economics of Education Review*, 48(2015): 86-101.
- Steinberg, M. P. y M. L. Donaldson. 2016. The New Educational Accountability: Understanding the Landscape of Teacher Evaluation in the Post-NCLB Era. *Education Finance and Policy*, 11(3): 340-359. Disponible en: https://doi.org/10.1162/EDFP_a_00186.
- Steinberg, M. P. y L. Sartain. 2015. Does Teacher Evaluation Improve School Performance? Experimental Evidence from Chicago's Excellence in Teaching Project. *Education Finance and Policy*, MIT Press, Vol. 10(4), pp. 535-572 (octubre).
- Sun, Yulan F. 2017. "La evaluación Docente en Chile: Trayectorias y Aprendizajes". Presentación en el Centro de Medición MIDE.
- Taut, S., E. Valencia, D. Palacios, M. V. Santelices, D. Jiménez y J. Manzi. 2016. Teacher performance and student learning: linking evidence from two national assessment programmes. *Assessment in Education: Principles, Policy & Practice*, 23: 53-74.
- Taylor, E. S. y J. H. Tyler. 2012. The Effect of Evaluation on Teacher Performance. *American Economic Review*, 102 (7): 3628-51.
- Unesco (Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura). 2015. Las carreras docentes en América Latina. La acción meritocrática para el desarrollo profesional. Santiago de Chile: UNESCO.
- Weisberg, D., S. Sexton, J. Mulhern, D. Keeling, J. Schunck, A. Palcisco y K. Morgan. 2009. The Widget Effect: Our National Failure to Acknowledge and Act on Differences in Teacher Effectiveness. Nueva York: The New Teacher Project (TNTTP).
- Wyckoff, J. y V. Katz. 2018. Review of Teacher Evaluation Systems in the United States: Technical Report and Comparative Analysis. Documento de antecedentes preparado para el proyecto del BID "Measuring teacher effectiveness in Latin America", 2017-18.

La evaluación docente es un instrumento fundamental para apoyar el desarrollo profesional de los maestros y mejorar la efectividad de la enseñanza. Este estudio analiza las evaluaciones docentes en América Latina y el Caribe y en varios sistemas educativos de buen desempeño de otras regiones, y presenta orientaciones prácticas sobre cómo desarrollar evaluaciones docentes exitosas.

Existen cinco claves para una evaluación docente exitosa. Primero, definir qué es lo que se espera de un docente de excelencia. Segundo, precisar cuál es el objetivo que busca alcanzar la evaluación docente. Tercero, producir resultados confiables y válidos. Cuarto, usar de manera eficiente los resultados según los objetivos establecidos. Quinto, realizar una investigación continua de los procesos y resultados para retroalimentar el sistema de evaluación.

Los países de América Latina y el Caribe aún tienen un largo camino que recorrer para implementar evaluaciones docentes sistémicas cuyos resultados se utilicen para mejorar la enseñanza. Este es un camino necesario para asegurar que todos los niños, niñas y jóvenes de la región tengan acceso a los mejores maestros.

El **Banco Interamericano de Desarrollo (BID)** es una institución internacional creada en 1959 para promover el desarrollo económico y social en América Latina.

