

Diseño de índices de calidad escolar: Lecciones de la experiencia internacional

Gregory Elacqua
Matias Martinez
Anne Sofie Welsth Olsen

División de Educación
Sector Social

NOTA TÉCNICA N°
IDB-TN-01706

Diseño de índices de calidad escolar: Lecciones de la experiencia internacional

Gregory Elacqua
Matias Martinez
Anne Sofie Welsth Olsen

Agosto 2019

Catalogación en la fuente proporcionada por la
Biblioteca Felipe Herrera del
Banco Interamericano de Desarrollo
Elacqua, Gregory M., 1972-
Diseño de índices de calidad escolar: lecciones de la experiencia internacional /
Gregory Elacqua, Matias Martinez von der Fecht, Anne Sofie Westh Olsen.
p. cm. — (Nota técnica del BID ; 1706)
Incluye referencias bibliográficas.
1. Educational indicators. 2. Educational accountability. 3. Educational evaluation. I.
Martinez von der Fecht, Matias. II. Olsen, Anne Sofie Westh. III. Banco
Interamericano de Desarrollo. División de Educación. IV. Título. V. Serie.
IDB-TN-1706

<http://www.iadb.org>

Copyright © 2019 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.



Diseño de índices de calidad escolar: Lecciones de la experiencia internacional

Gregory Elacqua, Matias Martinez von der Fecht & Anne Sofie Westh Olsen

División de Educación

Sector Social

Banco Interamericano de Desarrollo

RESUMEN

Durante los últimos años, varios países de América Latina y el Caribe (ALC) han fortalecido sus procesos de evaluación del aprendizaje y la calidad educativa. En este contexto, se vuelve un desafío encontrar un uso significativo a la extensa información recopilada que entregue retroalimentación efectiva a actores clave dentro del sistema escolar y que promueva una mejora continua del servicio que entregan las escuelas. Una estrategia comúnmente adoptada es la creación de índices de calidad escolar que midan de forma sintética y precisa la calidad del servicio. Con el objetivo de informar al diseño de índices de calidad escolar, en este documento describimos diversos sistemas de medición de la calidad de escuelas que se basan en distintas lógicas. Por un lado, revisamos las experiencias que representan una lógica más descentralizada en la recopilación de información, como en Hong Kong y Holanda. Estos casos muestran una concepción de calidad de escuelas más amplia que no se limita al establecimiento de metas desde el nivel central para resultados de aprendizaje, retención y desarrollo de habilidades, sino que también incentivan a las escuelas a implementar procesos de medición y evaluación de procesos educativos y elementos de la cultura educativa. Por el otro lado, examinamos sistemas de evaluación más centralizados como el de Estados Unidos y varios países de América Latina (Chile, Brasil y Colombia) que se han enfocado más en definir metas para diversos resultados que se esperan de las escuelas dejando a cada establecimiento decidir sobre sus procesos y el uso de recursos. Del mismo modo, discutimos sobre los diversos usos que se les asigna a los indicadores de la calidad, siendo uno de los principales la identificación de escuelas de bajo desempeño. La experiencia internacional entrega valiosas lecciones sobre qué indicadores de calidad elegir, la forma de combinarlos y muestra la necesidad de crear un sistema de medición y evaluación de la calidad escolar que conecte indicadores a consecuencias validadas por la comunidad escolar que sean percibidas como justas en el sentido que no se establecen castigos por resultados que están fuera del control de los establecimientos.

Palabras clave: Accountability escolar; índices de calidad escolar; América Latina y el Caribe

JEL: I28 (education and government policy); J18 (Public policy); H52 (National Government Expenditures and Education)

1. Introducción

La creación de índices de calidad escolar es una estrategia comúnmente adoptada en países con sistemas de información sobre el aprendizaje estudiantil. Su objetivo central es medir de forma sintética y precisa la calidad del servicio que entregan las escuelas y, al mismo tiempo, guiar prácticas de escuelas, distritos, directivos y docentes. Los índices de calidad escolar pueden ser una herramienta efectiva para identificar a las escuelas de mejor y peor desempeño sostenido en el tiempo, focalizar recursos de manera más eficiente, y entregar información relevante a la gestión y dirección escolar.

Para alcanzar estos objetivos, el diseño de estos índices debe buscar asegurar que esté basado en mediciones precisas y confiables desde el punto de vista estadístico y que sean válidas para la comunidad escolar. En este sentido, los índices deben ser informativos, con una interpretación simple y estar asociado a acciones concretas que pueden llevar a mejoras en la calidad educativa que se reflejan en el valor del indicador.

La revisión de la experiencia internacional es un ejercicio clave para contribuir en la producción de indicadores de calidad escolar que sean válidos, informativos y confiables. Durante las últimas décadas varios sistemas educativos han implementado diversas estrategias para medir la calidad de sus escuelas. Por un lado, algunos sistemas escolares (ej. Hong Kong), han optado por descentralizar el proceso de generación de indicadores volviendo a las escuelas responsables de la recopilación de datos, creación de indicadores, reflexión de cómo interpretar la información y formulación de estrategias para la mejora continua del servicio educativo. Por otro lado, otros sistemas (ej. Estados Unidos) se basan en procesos más centralizados en la generación y recopilación de datos que establecen metas de resultados para las escuelas cuyo logro es usualmente monitoreado desde el nivel central.

Este reporte, cuyo objetivo es informar el diseño de índices de calidad escolar en América Latina y Caribe (ALC), se organiza en cuatro secciones principales. Luego de esta introducción, en la segunda sección describimos brevemente sistemas de medición de la calidad escolar que usan lógicas más descentralizadas enfocándonos en sistemas representativos de esta estrategia de evaluación (Hong Kong y Holanda). Estos sistemas entregan mayor responsabilidad a las escuelas en la medición y evaluación de la calidad educativa. Si bien, esta no parece ser una estrategia común en países de ALC – probablemente porque en la mayoría de los casos los recursos y capacidades están más restringidos a nivel de escuelas – es interesante conocer formas alternativas a las que dominan en la región que podrían servir como referencia para el desarrollo de un marco conceptual que guíe en el futuro hacia un rol más activo de las escuelas en la medición de calidad.

Por otro lado, en la tercera parte revisamos sistemas de evaluación más centralizados que se basan principalmente en la definición de metas desde el nivel central para resultados educativos. Dentro de este grupo examinamos los casos de Estados Unidos, Chile, Brasil y Colombia. Para estos países, mostramos que existen varias etapas en el proceso de generación de indicadores de calidad que involucran decisiones que varían entre sistemas. Primero, describimos variadas formas usadas para medir la calidad educativa. El indicador más usado en estos sistemas escolares es el

aprendizaje de los estudiantes, pero la forma de medirlo varía en términos de las asignaturas y en si se consideran medidas en un punto del tiempo o medidas de valor agregado que toman en cuenta la trayectoria del aprendizaje de cada estudiante en el tiempo. Asimismo, es común que se consideren otras variables que miden otros resultados o dimensiones de la calidad del servicio tales como brechas de aprendizaje, tasas de graduación, asistencia a clases y habilidades en artes u otras áreas distintas a matemáticas, lenguaje y ciencias. Segundo, una vez que se eligen los indicadores, la definición de la importancia relativa que se le asigna a cada uno también varía según los objetivos de política educativa de cada sistema. Finalmente, mostramos diversas metodologías que se usan para combinar los distintos indicadores.

En la última sección presentamos recomendaciones para la construcción, diseño y uso de indicadores de calidad escolar que se basan en las experiencias de los sistemas educativos revisados. La experiencia internacional muestra que para un índice de calidad escolar sea un indicador informativo, válido y aceptado por la comunidad educativa se deben considerar varios factores: (i) la incorporación de múltiples dimensiones de calidad tales como resultados de aprendizaje, retención, desarrollo de habilidades sociales, medidas de equidad, entre otras; (ii) la elección y valoración de indicadores teniendo presente las diferencias de objetivos entre primaria y secundaria; (iii) factores técnicos asegurando una alta precisión estadística de los indicadores usados; (iv) inclusión de medidas de valor agregado que midan la trayectoria del aprendizaje de cada estudiante por varios años; (v) indicaciones claras y transparentes de cómo las escuelas pueden mejorar en cada dimensión medida; (vi) estrategias que involucren a los mismos profesionales de la educación en la recolección, uso y evaluación de datos relacionados a la calidad educativa; (vii) la discusión sobre las consecuencias asociadas a diversos valores del indicador debiera ser inclusiva y considerar la visión de distintos actores.

Adicionalmente, discutimos sobre distintos usos que se les asigna a los indicadores de la calidad a nivel de escuela. En algunos casos el uso para estos índices se ha enfocado en promover la innovación a través de la implementación de incentivos asociados a distintos valores de los indicadores de calidad. En otros casos, se ha usado para identificar escuelas de bajo desempeño y que buscar asegurar niveles mínimos de calidad de enseñanza por medio del establecimiento de consecuencias para las escuelas que muestran un bajo rendimiento de forma persistente.

En términos generales, las experiencias revisadas en este trabajo buscan entregar valiosas lecciones para alcanzar el mejor diseño posible de índices de calidad escolar, proveyendo antecedentes para apoyar decisiones clave en el proceso de generación de estos índices y, al mismo tiempo, evitar errores que otros sistemas han cometido.

2. Sistemas de medición de calidad escolar más descentralizados: Los casos de Hong Kong y Holanda

Los sistemas más descentralizados en la medición y evaluación de la calidad escolar buscan involucrar a los propios establecimientos educativos en el proceso. El caso más paradigmático es el de uno de los sistemas escolares de más alto rendimiento académico en el mundo: Hong Kong. Su Ministerio de Educación desarrolló un marco conceptual que define las principales dimensiones de calidad que las escuelas deben medir, evaluar y reportar. Otro caso interesante es el de Holanda, en donde se combinan procesos de medición de la calidad centralizados y descentralizados, pero con una tendencia en los últimos años hacia la descentralización.

2.1 Hong Kong

El sistema de medición de la calidad escolar en Hong Kong está basado en un marco conceptual elaborado en 2016. El sistema busca responsabilizar a las escuelas por la conducción de una autoevaluación holística basada en lineamientos generales que se enfoca en el autodesarrollo sostenible y permanente. Para lograrlo las escuelas deben seguir la lógica del ciclo virtuoso de “Planeamiento-Implementación-Evaluación” definido por las autoridades. El objetivo es que las escuelas desarrollen la capacidad de usar su propia experiencia implementando y evaluando sus políticas escolares para usarlas como un insumo clave para la planificación en el próximo ciclo educativo.

Para guiar y asegurar una autoevaluación de alta calidad el Ministerio de Educación creó un marco conceptual que operacionaliza el concepto de calidad escolar. Específicamente, definió cuatro dominios de calidad que se subdividen en ocho áreas clave, las cuales se subdividen en 23 indicadores de desempeño. Adicionalmente, pone a disposición de los establecimientos cinco herramientas que facilitan la implementación del proceso evaluativo: (i) definición de indicadores y medidas concretas asociadas a cada una de las 8 áreas de calidad escolar, las que sirven para definir una plataforma común entre las escuelas para su autoevaluación y reporte a las autoridades centrales; (ii) cuestionarios para que sean aplicados a profesores, especialistas, padres y estudiantes que buscan capturar la impresión de actores clave del proceso educativo sobre los distintos procesos educativos y, al mismo tiempo, facilitar la auto reflexión de sus actividades; (iii) sistema de información en línea que reporta de forma confidencial el valor agregado académico de la escuela y que la sitúa en una de tres categorías: debajo de lo esperado, similar a lo esperado o por sobre lo esperado. La idea de esta herramienta es aumentar la transparencia de las exámenes públicas y entregar información útil para el análisis y reflexión en cada establecimiento; (iv) instrumentos para evaluar el desarrollo afectivo y social de estudiantes de primaria y secundaria; (v) plataforma en línea para gestionar los datos recopilados para la autoevaluación y generar reportes para el Ministerio de Educación. Esta herramienta busca aliviar el trabajo administrativo en la escuela y facilitar la interpretación de la información de cada área de calidad.

Los dominios, áreas, indicadores y medidas específicas de desempeño se resumen en la Tabla 1. Tres de los dominios se asocian a procesos organizativos o pedagógicos: (i) administración y

organización; (ii) enseñanza y proceso de aprendizaje y; (iii) apoyo a estudiantes y ambiente escolar. El cuarto dominio se enfoca en resultados de desarrollo afectivo, social y académico, siendo este último componente el más importante. Entre los indicadores de calidad hay una combinación de indicadores objetivos, como el valor agregado de las escuelas o los resultados de desarrollo afectivo y emocional de los estudiantes, y también subjetivos, que provienen de los cuestionarios aplicados a profesores, especialistas, padres y estudiantes.

Tabla 1: Marco conceptual de la calidad de escuelas en Hong Kong

Dominio	Área	Indicador de desempeño	Medidas clave
I. Administración y Organización	1. Administración de la escuela	1.1 Planificación	i. Percepción de profesores, especialistas, padres y estudiantes sobre la (a) administración escolar, (b) liderazgo profesional y (c) desarrollo profesional docente. ii. Entrenamiento profesional en necesidades educativas especiales
		1.2 Implementación	
		1.3 Evaluación	
	2. Liderazgo profesional	2.1 Liderazgo y monitoreo	
		2.2 Colaboración y apoyo	
		2.3 Desarrollo profesional	
II. Aprendizaje e Enseñanza	3. Curriculum y evaluación	3.1 Curriculum y organización	i. Percepción de profesores, especialistas, padres y estudiantes sobre la (a) enseñanza, (b) aprendizaje de estudiantes, (c) curriculum y (d) evaluación del curriculum. ii. Porcentaje de lecciones en áreas clave de aprendizaje iii. Número de días con actividades escolares iv. Asignaturas disponibles para secundaria
		3.2 Implementación del curriculum	
		3.3 Evaluación del desempeño	
		3.4 Evaluación del curriculum	
	4. Aprendizaje de estudiantes y enseñanza	4.1 Proceso de aprendizaje	
		4.2 Desempeño en aprendizaje	
		4.3 Organización de la enseñanza	
		4.4 Proceso de la enseñanza	
		4.5 Retroalimentación y seguimiento	
III. Apoyo a Estudiantes y Ambiente Escolar	5. Apoyo a estudiantes	5.1 Apoyo para el desarrollo de estudiantes	i. Percepción de profesores, especialistas, padres y estudiantes sobre (a) el apoyo para el desarrollo de estudiantes, (b) cooperación entre hogar y escuela y (c) clima escolar ii. Destino de los graduados
		5.2 Clima escolar	
	6. Asociaciones	6.1 Cooperación entre el hogar y la escuela	
		6.2 Vínculos con organizaciones externas	
IV. Desempeño de los estudiantes	7. Actitud y comportamiento	7.1 Desarrollo afectivo	i. Indicadores de estudiantes: (a) tasa de asistencia, (b) actitud hacia la escuela, (c) resultados en examinación pública, (d) peso corporal aceptable, (e) participación en competiciones escolares y comunidad ii. Valor agregado académico de la escuela
		7.2 Desarrollo social	
	8. Participación y logro	8.1 Desempeño académico	
		8.2 Desempeño no académico	

2.2 Holanda

Durante el 2008, Holanda comenzó a usar un sistema de inspección de escuelas cuyo objetivo es medir el riesgo que presentan de entregar una educación de baja calidad. Este sistema se compone de dos etapas principales. En la primera el Inspectorado de la educación escolar usa información administrativa para calcular indicadores de alerta temprana para identificar a un grupo de escuelas que potencialmente pueden ser consideradas como de baja calidad. En la segunda, se establece un monitoreo más intenso de las escuelas dependiendo de sus resultados de la primera etapa, y se acuerda una planificación para mejorar los resultados en los casos de resultados más bajos. Si luego de dos años, no hay mejora, las escuelas pueden enfrentar sanciones administrativas y/o financieras.

Durante la primera de estas etapas, el Inspectorado analiza la calidad escolar usando los resultados de estudiantes en pruebas estandarizadas de todas las asignaturas medidas los últimos tres años, información del contexto socioeconómico de las familias y datos del desempeño anual de los establecimientos, tales como rotación del personal escolar, número de estudiantes y gestión

financiera. Los resultados de aprendizaje para cada nivel socioeconómico de las familias es el indicador principal del análisis que se usa para clasificar a las escuelas en alguna de las siguientes tres categorías: (i) categoría verde que se usa para las escuelas que “no tienen riesgo de fallar;” (ii) naranja, utilizado para identificar a las escuelas que tienen un “riesgo potencial de fallar;” (iii) roja, la que agrupa a las escuelas que tienen un “alto riesgo de fallar.”

Durante la segunda etapa, las escuelas en la categoría naranja son analizadas en más detalle por el Inspectorado. Se les solicita a las mismas escuelas por más información sobre su desempeño y resultados de aprendizaje en otros grados que no son medidos por la evaluación estandarizada nacional. Si el Inspectorado determina que los resultados son suficientemente buenos, la escuela es reclasificada en la categoría verde, pero se le da una advertencia informal si sus resultados de aprendizajes están por debajo del promedio nacional o si han venido decreciendo en el tiempo. Si, luego de analizar la información adicional, el Inspectorado considera que existen riesgos potenciales de fallar, se coordina una reunión con la dirección de la escuela. En la reunión se discute sobre estos riesgos. Si la escuela no provee suficiente información para convencer al Inspectorado de que tienen la capacidad de hacer frente a los riesgos potenciales, entonces se establece una visita adicional a la escuela. Durante la visita se entrevista al personal escolar y se evalúan las prácticas de la escuela. Dependiendo de la evaluación durante la visita, la escuela es clasificada como de (i) calidad suficiente, (ii) en riesgo de fallar; (iii) altamente subdesarrollada.

En el caso de las escuelas que fueron clasificadas durante la primera etapa en la categoría roja, el Inspectorado también realiza un análisis más detallado de sus datos y coordina una reunión con la dirección del establecimiento. Luego de la inspección adicional, estas escuelas se clasifican como (i) en riesgo de fallar o (ii) altamente subdesarrolladas. A ambos grupos de escuelas, el Inspectorado les solicita la elaboración de un plan de mejoramiento de la calidad, el cual debe definir indicadores que se examinarán en futuras visitas de inspección. Si una escuela en la categoría “en riesgo de fallar” no mejora al cabo de dos años, se reclasifica como “altamente subdesarrollada.” Los establecimientos en esta última categoría se enfrentan a más actividades de inspección, tales como reuniones entre el Inspectorado y la dirección de la escuela o advertencias oficiales. Si estas actividades no logran mejorar la calidad de estas escuelas, el Inspectorado reporta a la escuela “altamente subdesarrollada” al Ministerio de Educación proponiendo diversos tipos de sanciones administrativas y/o financieras. Basado en este reporte, el Ministerio de Educación decide si concreta o no las sanciones propuestas [1].

En los últimos 10 años el sistema se ha ido actualizando. En 2017 se definió un nuevo marco de evaluación para cada nivel educativo. Este marco define los factores de una educación de calidad y los estándares básicos asociados a cada una de las dimensiones, es decir, las prácticas y resultados para considerar al servicio como de calidad mínimamente aceptable. Adicionalmente, define otro conjunto de estándares asociados a un servicio de alta calidad. Este marco se entrega a las escuelas para informarles sobre qué es lo que se espera de su trabajo.

Para el caso de primaria, se definen 17 factores asociados a un servicio de calidad que se agrupan en cinco dimensiones: (i) procesos educativos; (ii) clima escolar; (iii) resultados de aprendizaje; (iv) aseguramiento de la calidad y ambición; y (v) administración financiera. La Tabla 2 presenta una breve descripción de cada uno de los factores.

Para cada uno de los factores se definen estándares de calidad básica y de alta calidad de manera de guiar las prácticas de las escuelas. Por ejemplo, en el caso del factor de “didáctica” dentro del grupo de procesos educativos, el estándar básico corresponde al cumplimiento de tres condiciones: (i) la existencia de un programa de planificación para cada clase; (ii) unidades de clases que están estructuradas lógicamente; (iii) las clases se diseñan según necesidades de los estudiantes. Para alcanzar el estándar de alta calidad las escuelas deben cumplir con las condiciones anteriores y con una cuarta condición: (iv) los docentes entregan retroalimentación relevante a los estudiantes y usan el tiempo de clase de forma eficiente [2].

Tabla 2: Factores asociados a la calidad de escuelas de primaria en Holanda

Proceso educativo
<i>Provisión</i> : Escuela prepara a estudiantes para el futuro y vida en sociedad
<i>Desarrollo de perspectivas</i> : Escuela tiene sistema de monitoreo del desarrollo de los estudiantes
<i>Didáctica</i> : Profesores estructuran actividades de manera que permite el desarrollo de estudiantes
<i>Apoyo adicional</i> : Los estudiantes que lo necesitan reciben más apoyo, supervisión e instrucción
<i>Cooperación</i> : Escuela tiene convenios y socios relevantes en la provisión del servicio educativo
<i>Evaluación y progreso</i> : Evaluación y guía hacia el siguiente nivel escolar se realiza con cuidado
Clima escolar
<i>Seguridad</i> : Dirección y profesores proveen un clima de aprendizaje seguro
<i>Clima de enseñanza</i> : El ambiente escolar asegura apoyo constante
Resultados de aprendizaje
<i>Resultados</i> : Cumple con mínimos requeridos
<i>Habilidades sociales</i> : Estudiantes aprenden estas habilidades
<i>Logro en niveles futuros</i> : Resultados de estudiantes en siguiente nivel educativo
Aseguramiento de la calidad
<i>Aseguramiento de la calidad</i> : Escuela ha instalado un sistema de aseguramiento de la calidad
<i>Cultura de calidad</i> : Operaciones transparentes y orientadas al servicio de calidad
<i>Accountability y diálogo</i> : Escuelas enfrentan accountability interno y externo. Existen instancias de diálogo
Manejo financiero
<i>Continuidad</i> : Escuela puede cumplir sus obligaciones financieras en el corto y largo plazo
<i>Eficiencia</i> : Escuela hace un uso eficiente y efectivo de sus recursos financieros
<i>Legitimidad</i> : El manejo financiero está alineado con la legislación

3. Sistemas de medición de calidad más centralizados: Estados Unidos y América Latina

Los sistemas más centralizados se basan en el establecimiento de metas definidas desde el nivel central. El rol de las escuelas en el levantamiento de información es mínimo o no existe. En esta sección revisamos los casos de Estados Unidos y de algunos países de ALC donde está lógica domina. La revisión de la experiencia en estos países entrega algunas lecciones interesantes. Por ejemplo, el indicador más usado es el aprendizaje de los estudiantes que proviene de evaluaciones estandarizadas, pero la forma de medirlo es variada. Asimismo, es común que se consideren otras

variables que miden otros resultados o dimensiones de la calidad del servicio tales como brechas de aprendizaje, tasas de graduación, asistencia a clases y habilidades en artes u otras áreas distintas a matemáticas, lenguaje y ciencias. La definición de la importancia que se asigna a cada indicador también es variable y está fuertemente vinculada al nivel educativo y los objetivos de política escolar en cada sistema. Por último, las metodologías que se usan para combinar los distintos indicadores también presentan una alta variabilidad.

3.1 Estados Unidos

El caso de Estados Unidos es interesante pues luego de que la política *No Child Left Behind* (NCLB) se transformara en ley federal el 2002, se les exigió a todos los estados evaluar los aprendizajes de sus estudiantes entre tercero y octavo grado, y en otro grado de secundaria. Al mismo tiempo, se estableció que las escuelas fueran evaluadas según si sus estudiantes (y algunos subgrupos como estudiantes de raza negra o latinos) lograban un progreso adecuado de un año al siguiente. Si las escuelas o distritos escolares no lograban un progreso adecuado se deberían determinar consecuencias o sanciones. Luego de un poco más de 10 años de su implementación, muchos estados comenzaron a presionar para usar indicadores más diversos para medir la calidad de las escuelas. Ante esto, la política fue reemplazada en 2015 por la ley *Every Student Succeeds* (ESSA) la cual permitió la inclusión de diferentes indicadores de calidad de las escuelas según las preferencias de cada estado.

Lo anterior derivó en el establecimiento de diversos indicadores y metodologías para medir la calidad y desempeño de escuelas y distritos escolares. De los 50 estados y el Distrito de Columbia (D.C.) en Estados Unidos, 48 estados y D.C. han instituido un único (o incluso múltiples) índice(s) de desempeño para medir el funcionamiento de escuelas y distritos.¹

Varios estudios recientes se han centrado en analizar la estructura subyacente de los indicadores construidos, identificando y comparando las medidas usadas, las metodologías seguidas y los usos que se les ha dado [3, 4, 5]. Las decisiones sobre qué variables incluir, cómo definir las, combinarlas y la elección de valores que se pueden interpretar como adecuados tienen consecuencias relevantes en los incentivos que se generan. Por ejemplo, durante el primer año en que NCLB entró en vigencia, el porcentaje de escuelas clasificadas como de baja calidad en distintos estados de Estados Unidos presentó un rango que iba desde el 1% al 80% [5].

En general, una de las primeras decisiones clave es la definición de los indicadores que se considerarán en la generación del índice de calidad escolar. La Tabla 3 fue extraída de uno de los estudios que analizan los 51 sistemas de *accountability* en Estados Unidos a un año de la aprobación de ESSA. Los datos muestran un resumen de los indicadores usados en los 48 estados y D.C. que calculan índices de desempeño de escuelas y distritos. En el estudio, los indicadores

¹ Sólo dos estados, Vermont y Montana, usan una serie de indicadores de forma individual para medir el progreso de sus estudiantes. Por ejemplo, Vermont clasifica el desempeño de escuelas y distritos como satisfactorio basado en indicadores individuales de resultados de pruebas estandarizadas de matemáticas y lectura, tasa de participación en estas evaluaciones y tasa de graduación.

fueron agrupados en 14 categorías. Ningún estado usa estas 14 categorías simultáneamente, y tan sólo tres usan un máximo de 10 indicadores de forma simultánea. Por otro lado, sólo 3 estados usan tres o menos indicadores. El número de indicadores que se combinan para generar el índice de calidad que más se repite es 8 (en 15 estados).

Tabla 3: Indicadores considerados en el cálculo de índices de desempeño en 48 estados de Estados Unidos y D.C.

Elemento	Estados	
	Número	Porcentaje
Prueba estandarizada	49	100%
Indicador de crecimiento de aprendizajes	46	94%
Tasa de graduación	45	92%
Cierre de la brecha de aprendizaje ^(a)	42	86%
Participación en pruebas estandarizadas ^(b)	32	65%
Grado de preparación post escuela ^(c)	30	61%
Grado de preparación para una carrera ^(d)	27	57%
Tasa de asistencia	20	41%
Peso por rendimiento ^(e)	13	29%
Tasa de deserción	10	22%
Efectividad del educador y ambiente escolar ^(f)	8	16%
Peso por materia evaluada	6	14%
Indicador de dominio del inglés	6	12%
Artes o humanidades	5	10%

(a): Por ejemplo, en Pennsylvania se define la brecha de aprendizaje en una línea entre estudiantes avanzados y otros grupos vulnerables de estudiantes y luego se dan puntos a quienes logran avanzar en el cierre de estas brechas

(b): Muchos estados castigan a las escuelas en donde la tasa de participación en pruebas estandarizadas es menor al 95%

(c): Entre las variables que se usan están el promedio en pruebas estandarizadas de admisión a la universidad, porcentaje de estudiantes que toman clases avanzadas y tasas de admisión a la universidad

(d): Varios estados han desarrollado tests específicos para medir esta variable

(e): En algunos casos los estudiantes de bajo y alto rendimiento no pesan lo mismo

(f): Se usan variables como porcentaje de docentes de alta calificación o grado académico. Del mismo modo, se usan indicadores de expectativas, comunicación, seguridad y respeto en la escuela

Respecto al tipo de indicador usado, sobre el 90% de los estados basan sus indicadores en los resultados de pruebas de aprendizaje, su evolución en el tiempo y la tasa de graduación. De forma similar, un alto porcentaje de estados incluye como indicador la capacidad para cerrar brecha de aprendizaje entre distintos grupos de estudiantes. Por ejemplo, en Illinois se asigna puntos basado en si es que la escuela logra metas de reducción de brechas de aprendizaje para grupos de estudiantes que históricamente han mostrado un bajo desempeño, entre los cuales se encuentran

minorías raciales, aquellos en desventaja económica y para quienes el inglés no es su lengua materna.

Cada indicador adopta distintas formas. Para el caso específico de la medición del aprendizaje, la mayoría de los estados sólo usan los resultados de matemáticas y lenguaje (ej. Nevada), sin embargo, algunos otros usan el desempeño en matemáticas, lectura, escritura, ciencias y ciencias sociales (ej. California). Incluso, en algunos casos, se les asigna un peso diferenciado a estudiantes de alto y bajo rendimiento (ej. en Ohio se usa un multiplicador de 1.2 al porcentaje de estudiantes en nivel avanzado en la prueba estandarizada del estado y de 0.3 a los que están en el nivel bajo) o al resultado de distintas asignaturas (ej. Oregon).

Asimismo, cuando se incluyen medidas de mejoras en el aprendizaje se siguen principalmente tres metodologías: (i) progreso anual del distrito o escuela; (ii) progreso anual de estudiantes en comparación a pares con resultados similares en la línea de base; (iii) modelos de valor agregado en que se calcula el valor de mejoras esperado para cada escuela y se compara con la mejora que tuvo en realidad.

Una vez definidas las variables, la forma de combinarlas también implica adoptar distintas estrategias. Por ejemplo, sólo tomando en cuenta la forma en que los aprendizajes son incorporados en la construcción de indicadores globales hay una amplia diversidad de métodos. La mayoría de los estados que combinan resultados de varios niveles dándole un peso mayor a los aprendizajes en educación básica en comparación con educación secundaria. No obstante, el peso específico varía, así como también si se considera sólo el nivel o mejoras de un año al siguiente.

Para tener una impresión más precisa de cómo se combinan los distintos indicadores, el Centro para Progreso Americano clasificó los indicadores que se usan en los 51 sistemas de *accountability* a un año de la implementación de ESSA y los agrupó en siete categorías. Luego, para 36 de los 51 sistemas logró identificar la importancia relativa de cada indicador en el índice final que mide la calidad de las escuelas en educación primaria/intermedia y secundaria. La Tabla 4 muestra un resumen de los indicadores usados, el número de estados que los incluye en sus índices de calidad y el peso promedio que le asignan en cada nivel educativo.

El indicador más importante en los sistemas de *accountability* en Estados Unidos es el aprendizaje. En conjunto, el logro académico y el progreso en el aprendizaje de un año al siguiente representan el 91% del indicador de calidad de escuelas en primaria/intermedia, y un 64% en secundaria. La menor ponderación en secundaria se debe a que muchos estados valoran más la capacidad de la escuela para retener estudiantes en este nivel. Por ejemplo, en los estados de Illinois, Oregon, Washington y Wisconsin se pondera con más de un 30% a las medidas de persistencia o permanencia estudiantil. Adicionalmente, muchos estados agregan indicadores que muestren la capacidad de las escuelas para preparar a sus estudiantes para la educación superior. En este sentido, medidas de rendimiento en pruebas de admisión a universidades y la proporción de estudiantes en clases avanzadas se vuelven más importantes. En Alabama, Arkansas, Indiana y Dakota del sur la ponderación de esta dimensión alcanza un 30%.

Tabla 4: Pesos asociados a las variables consideradas en el cálculo de indicadores de calidad

Elemento	Número de estados	Peso promedio del indicador ^(a)	
		básica/intermedia	secundaria
Nivel de aprendizaje	51	51%	42%
Persistencia ^(b)	51	3%	22%
Mejoras de aprendizaje	46 ^(c)	45%	30%
Preparación para la educación superior ^(d)	30	11% ^(e)	15%
Alerta temprana ^(f)	24	11%	7%
Adquisición de inglés	6	7%	6%
Otros ^(g)	27	10%	12%

(a): El indicador del peso de cada indicador está basado en datos de 36 estados

(b): La persistencia escolar se mide principalmente por las tasas de graduación y tasas de deserción.

(c): En básica/intermedia son 32 estados y en secundaria 25

(d): Incluye variables como participación en clases avanzadas durante la secundaria y puntajes en pruebas estandarizadas de admisión a la universidad

(e): Este indicador no es usado para evaluar la calidad en el nivel básico

(f): La alerta temprana considera tres indicadores: tasa de asistencia (18 estados), ausentismo crónico (5) y un indicador de si los estudiantes asisten al grado correspondiente a una graduación a tiempo (5)

(g): Algunos indicadores en esta categoría son mejoras en lectura de K-3, porcentaje de profesores que usan datos longitudinales del sistema, retención de docentes, clima escolar y habilidades

En el caso de la educación primaria, más allá de la alta ponderación que se da a los indicadores relacionados a las pruebas estandarizadas, existe una alta variación entre estados en la forma en que se combinan los indicadores disponibles. En Louisiana se da un peso de 100% al nivel de logro en las pruebas estandarizadas y 0% a los avances de este puntaje. Por otro lado, Idaho y Oregon dan una ponderación de 25% al puntaje en un año específico y un 75% a los avances de un año al siguiente.

Adicionalmente, también es común el uso de otros indicadores menos convencionales, que en general, reflejan la diversidad de preferencias de las autoridades en cada sistema y los factores que consideran que el sistema escolar debería desarrollar. Por ejemplo, Ohio incluye indicadores de alfabetización temprana para los niveles entre kínder y tercero básico. Georgia considera el porcentaje de profesores que hacen uso del sistema de información longitudinal de estudiantes y el grado de integración de estudiantes con discapacidades en clases regulares. Iowa pondera la retención de docentes en cada escuela. Illinois añade indicadores de clima escolar. Nuevo México incluye datos de su encuesta a estudiantes sobre su percepción de la calidad de la instrucción que reciben. Finalmente, California, además de considerar datos de la percepción de profesores, estudiantes y padres sobre la cultura escolar, agrega indicadores de habilidades socio emocionales de estudiantes.

3.2 América Latina

América Latina no se caracteriza por contar con sistemas robustos para la medición de la eficiencia y calidad escolar: (i) Ningún sistema de la región cuenta con mediciones sistemáticas de aprendizaje en años contiguos que permitan la estimación de modelos de valor agregado en múltiples niveles; (ii) no se mide el nivel de vocabulario antes de entrar a la escuela a pesar de la existencia de instrumentos para ello; (iii) la mayoría de los países no cuentan con estadísticas confiables del gasto por estudiante en cada escuela que midan el nivel de recursos con los que cuentan para alcanzar distintos niveles de aprendizaje; y (iv) en casi ningún sistema se han definido estándares para prácticas organizacionales o pedagógicas de alta calidad, ni criterios que guíen los procesos dentro de las escuelas.

A pesar de lo anterior, algunos países de la región han comenzado a invertir sistemáticamente en consolidar sistemas de información sobre la calidad escolar. En general, estos sistemas recopilan información sobre resultados de aprendizaje en algunos grados, retención de estudiantes, y otras medidas de eficiencia escolar. Por ejemplo, en Panamá, desde 2016 se ha comenzado a implementar un plan de evaluación educativa que mide el desempeño de centros educativos, docentes y estudiantes, incluyendo una evaluación censal del aprendizaje de estudiantes de tercer grado en lectoescritura, matemáticas y ciencias. En Ecuador, desde el año 2013 se mide regularmente el aprendizaje de estudiantes de varios grados en distintas materias y desde 2016 se ha comenzado a realizar una evaluación censal de docentes que mide su preparación académica para realizar clases. En Paraguay, en 2015 se implementó por primera vez una evaluación censal que midió el aprendizaje de todos los estudiantes en los grados 3, 6, 9 y 12 en el área de matemática, castellano y guaraní. En los casos más sofisticados de la región se han elaborado indicadores que contribuyen a la identificación de escuelas de peor calidad para la focalización de programas o para incentivar a las escuelas a mejorar su desempeño.

En esta sección revisamos los sistemas de Chile, Brasil y Colombia pues tienen una mayor experiencia diseñando y calculando indicadores de calidad a nivel de escuelas que han sido asociados a incentivos que buscan modificar sus prácticas organizacionales y pedagógicas de manera tal que las lleve a mejorar su desempeño.

Chile I: Sistema Nacional de Evaluación del Desempeño de los Establecimientos Educativos Subvencionados (SNED)

En 1995 se estableció un sistema para la evaluación del desempeño de las escuelas para incentivar mejoras en sus prácticas a través de la entrega un incentivo monetario a los profesionales que trabajan en las escuelas mejor evaluadas. Para evaluar a los establecimientos, SNED calcula un índice basado en 6 dimensiones: (i) efectividad, medida por el nivel de aprendizaje considerando las características de los estudiantes; (ii) superación, que corresponde a cambios del logro educativo en la escuela durante de un año al siguiente; (iii) iniciativa, que muestra la capacidad del establecimiento para incorporar innovaciones educativas; (iv) mejoramiento de las condiciones de trabajo y adecuado funcionamiento de la escuela; (v) igualdad de oportunidades, entendida como

la capacidad de la escuela para dar acceso y asegurar la permanencia de los estudiantes en el sistema escolar, y para integrar estudiantes con dificultades de aprendizaje; (vi) integración y participación de docentes y padres en el desarrollo del proyecto educativo del sistema.

Una característica interesante del diseño de este programa es que antes de operacionalizar cada dimensión, se generaron 125 grupos de escuelas para intentar asegurar la comparación entre escuelas que enfrentan un contexto similar. Para ello, se usa un análisis de conglomerados y tres variables para agrupar a las escuelas que más se parecen entre sí en relación con (i) la zona geográfica en que se ubican, urbana o rural; (ii) nivel de educación que imparten, básica, media o ambos; (iii) nivel socioeconómico, medido por un índice de vulnerabilidad del establecimiento, el ingreso del hogar de los estudiantes y la escolaridad de sus padres. Luego, cada dimensión del índice y el índice final se compara sólo entre escuelas dentro de un mismo grupo.

La primera dimensión de efectividad tiene un peso en el indicador final de 37% y corresponde al promedio de los puntajes de matemáticas, lenguaje, ciencias y ciencias sociales para cuarto, octavo y décimo grado. En el caso de cuarto se usan las últimas dos mediciones disponibles. Para octavo y décimo grado se usa sólo una medición. Todos estos puntajes provienen de los datos levantados en el Sistema de Medición de la Calidad de la Educación (SIMCE).²

La medida de superación se pondera con el 28% y se define como la diferencia en el puntaje de las pruebas SIMCE entre dos años consecutivos. Se usan los mismos grados y asignaturas que para la medida de efectividad. Las diferencias en cada grado y prueba se corrigen según su significancia estadística. En particular, las diferencias corregidas son iguales a las diferencias de los puntajes estandarizados para los casos en que éstas sean significativas estadísticamente e iguales a 0 en los casos en que no son significativas estadísticamente. El factor final considera el promedio de estas diferencias corregidas de cada nivel, de acuerdo con la información existente para cada establecimiento.

Para medir la iniciativa de la escuela, se usan 11 indicadores que provienen de una encuesta a directores que recopila información sobre la frecuencia con que se realizan actividades tales como: apoyo a prácticas profesionales, planificación para capacitación docente, actividades extracurriculares de libre elección e inter escolares, entre otras. Asimismo, se incluye como indicador la existencia de un equipo de gestión y consejo escolar.

La variable de mejoramiento de las condiciones de trabajo pesa 2% y está compuesta por dos tipos de indicadores: (i) variación en el número de sanciones leves y graves entre los dos últimos años y (ii) cumplimiento con los procesos de levantamiento de datos del censo docente, notas y promoción.

El componente de igualdad de oportunidades pesa un 22% y se divide en tres factores: (i) rendimiento escolar medido por la tasa de aprobación y retención de los estudiantes; (ii) integración escolar correspondiente al porcentaje de estudiantes con déficits de aprendizaje que son integrados al sistema escolar regular, y a un indicador que muestra si la escuela desarrolla un

² El Sistema de Medición de la Calidad de la Educación o SIMCE corresponde al sistema usado en Chile para la evaluación de aprendizajes. Se creó en 1988 y en la actualidad recopila información sobre rendimiento en pruebas estandarizadas para los grados 2, 4, 6, 8 y 10, además de datos sobre el contexto de enseñanza.

proyecto de integración escolar; (iii) dimensión que mide la ausencia de prácticas discriminatoras y sanciones indebidas sobre alumnos y alumnas.

La última dimensión de integración de padres y apoderados tiene un peso de 5% y usa información proveniente de cuestionarios a padres y directores de escuela. Se considera si existe un consejo de profesores, centro de padres, centro de estudiantes, si se incorpora a la comunidad en compromisos educativos, si la escuela informa los resultados del SIMCE y del SNED a las familias y si el establecimiento integra a padres en las decisiones que toma. Todos estos indicadores tienen una misma ponderación para el cálculo de este componente.

Finalmente, se calcula el índice final usando un promedio ponderado y se entregan premios al 35% de establecimientos de mejor desempeño. El 25% superior recibe la subvención por desempeño de excelencia y los colegios entre el 26 y 35% superior de rendimiento reciben un 60% de la subvención de excelencia.

Chile II: Ley Subvención de Escolar Preferencial (SEP)

En 2008 se introdujo la ley SEP que aumentó el monto de recursos por estudiante transferido desde el gobierno central a las escuelas en cerca de un 50% para aquellos clasificados como estudiantes vulnerables. En total, estos recursos representan cerca del 16% del total de estas transferencias [6]. Junto a los recursos adicionales se introdujeron mecanismos de *accountability* que buscaban asegurar estándares mínimos de aprendizaje. Específicamente, para que la transferencia de recursos asociados a SEP se efectuara, las escuelas debían firmar el Convenio de Igualdad de Oportunidades y Excelencia Educativa con el Ministerio de Educación en el que se comprometen a documentar el uso que se da a los recursos, establecer metas de rendimiento académico para sus estudiantes y entregar información a los padres sobre el desempeño de la escuela.

Adicionalmente, con el objetivo de informar a la comunidad sobre el desempeño de las escuelas y asegurar una calidad mínima, el sistema clasificaba a las escuelas que participan de SEP según su desempeño. Hasta 2016, se consideraron tres categorías: (i) autónomas; aquellas que muestran un rendimiento sistemáticamente más alto que el promedio nacional, (ii) emergentes; escuelas que no tienen un desempeño sistemáticamente más alto al promedio; y (iii) en recuperación; establecimientos que tienen un rendimiento sistemáticamente más bajo al promedio.

La ley SEP obliga a todas las escuelas a elaborar un plan de mejoramiento que defina actividades y fije metas por los siguientes cuatro años. En el caso particular de las escuelas en recuperación, todos los fondos adicionales provenientes de SEP debían ser usados en la implementación de su plan de mejoramiento. Al mismo tiempo, si su categoría no mejoraba en tres años, el Ministerio de Educación enviaba información a las familias en el establecimiento incentivándoles a considerar otras opciones escolares para sus hijos. Si éstas decidían cambiar de escuela, el gobierno debía financiar su transporte. Aún más, si la escuela se mantenía en la categoría en recuperación por cuatro años consecutivos, el ministerio dejaba de entregarle financiamiento, lo que en términos prácticos implica el cierre del establecimiento.

Para que una escuela fuese clasificada en la categoría en recuperación debían cumplir con una serie de condiciones. Primero, se definían dos condiciones esenciales para que la escuela fuese clasificable: (i) que la escuela hubiese participado en al menos dos de las últimas tres mediciones disponibles del SIMCE para cuarto grado; (ii) que el número promedio de estudiantes que participa cada año en el SIMCE de lenguaje, matemáticas y ciencias para cuarto grado sea mayor a 20 estudiantes. El objetivo de estas condiciones es reducir errores de medida inherentes a una baja cantidad de estudiantes evaluados.

Segundo, se definieron otras dos condiciones que identificaban un bajo rendimiento en las pruebas estandarizadas de matemáticas, lenguaje y ciencias de cuarto grado: (i) teniendo en consideración que el puntaje de cada asignatura sigue una distribución normal con un promedio de 250 y desviación estándar de 50, se definió como una primera condición que el promedio de las tres asignaturas debe ser inferior a 220. La segunda condición (ii) es que la proporción de estudiantes con un puntaje promedio sobre 250 en ese mismo puntaje fuese inferior al 20%. Cuando ambas condiciones se cumplían en al menos dos de las últimas tres mediciones disponibles, la escuela era clasificada en la categoría en recuperación.

Un grupo adicional de escuelas también se clasificaban en recuperación, pero siguiendo otro criterio: estar por debajo del percentil 10 cuando las escuelas son ordenadas según el Índice de Calidad Educativa (ICE). El ICE es un índice compuesto por la suma ponderada del (i) valor estandarizado del promedio de puntajes SIMCE de matemáticas, lenguaje y ciencias de cuarto grado (ponderación del 70%); y (ii) el valor estandarizado del Índice de Indicadores Complementarios (ICC – 30% restante).

El ICC es otro promedio ponderado de los siguientes indicadores estandarizados: (i) tasa de aprobación de alumnos (25%); (ii) tasa de retención de estudiantes (25%); (iii) indicador que mide el grado de mejoramiento de condiciones de trabajo y qué tan adecuado es el funcionamiento en la escuela (17%); (iv) medida de la iniciativa y capacidad del establecimiento para incorporar innovaciones educativas y comprometer a agentes externos en su quehacer pedagógico (13%); (v) grado de integración de profesores y familias en el proyecto educativo de la escuela (13%); (vi) evaluación del cuerpo docente (7%). En el caso de las escuelas que no participan en la evaluación docente (sólo las públicas lo hacen), se usan el resto de las variables cambiando su ponderación a 25%, 25%, 20%, 15% y 15% respectivamente para los factores de (i) a (v).

Es importante mencionar que la primera vez que se usó esta metodología para identificar escuelas de bajo rendimiento fue en 2012 y que se dejó de usar en 2016. Un problema de diseño de esta clasificación fue que, en algunos casos, el progreso potencial que una escuela podía alcanzar, no se veía reflejado en su categoría al año siguiente. Por ejemplo, una escuela era clasificada en recuperación en 2012 según sus últimos tres resultados SIMCE disponibles: 2008, 2009 y 2010. Al año siguiente, en 2013, era clasificada nuevamente, pero con base a sus resultados 2009, 2010 y 2011, es decir, sin tomar en cuenta las mejoras que podrían haber logrado por cambios introducidos durante el 2012. De esta forma, para el 2015, una escuela podría ser clasificada en recuperación por cuatro años consecutivos basándose en sus resultados 2011, 2012 y 2013. En otras palabras, el desempeño del 2011, que es previo a la introducción de la categoría en

recuperación y a la elaboración e implementación del plan de mejoramiento que SEP requería, influía en la probabilidad de que una escuela perdiera su financiamiento público.

Otro problema fue que la clasificación de escuelas sólo se centraba sólo en los resultados de cuarto básico y en medidas del nivel de aprendizaje. Durante 2014, se aprobaron modificaciones a la legislación estableciendo que se debería buscar una nueva metodología para clasificar y ordenar escuelas. En particular, se definió que el nuevo ordenamiento de escuelas debía considerar todas las áreas evaluadas censalmente, incluir estándares de aprendizaje y grado de cumplimiento de otros indicadores de calidad educativa. También se deberían considerar características del establecimiento educacional, tales como su vulnerabilidad, y, cuando sea posible, usar indicadores de progreso o valor agregado. La nueva metodología comenzó a implementarse en 2016 y sus detalles se describen a continuación.

Chile III: Ley de Aseguramiento de la Calidad (SAC)

La nueva metodología para la nueva ordenación de escuelas según su calidad se describe en el decreto 17 que aprobaba cambios a la Ley 20.529. En este decreto se definen los objetivos generales de la ordenación, los requisitos para los indicadores usados, sus definiciones, la metodología para operacionalizar los indicadores y cómo son combinados entre sí. El objetivo principal de la ordenación es responsabilizar a las escuelas y su dirección por el cumplimiento de los estándares de aprendizaje y la calidad del servicio que ofrecen. Asimismo, busca identificar las necesidades de apoyo focalizado cuando corresponda. Para alcanzar estos objetivos, y asegurar la construcción de un indicador informativo, establece que la ordenación debe hacerse de manera anual y obligatoria, de forma independiente para la educación primaria y secundaria, definiendo cuatro categorías de desempeño: (i) alto; (ii) medio; (iii) medio-alto; (iv) insuficiente. Adicionalmente, la legislación estipula que la metodología debe actualizarse en un plazo no menor a cuatro años y que no exceda los ocho.

Para asegurar el uso de indicadores relevantes se establecen requisitos específicos. En primer lugar, se deben usar todos los resultados de aprendizaje disponibles que se han levantado de forma censal. Segundo, se deben vincular los resultados a estándares de aprendizaje que se definen en función del grado de cumplimiento del currículo escolar. Tercero, se deben incluir otros indicadores de calidad escolar y con sus respectivos estándares que permitan evaluar las acciones de los establecimientos. Cuarto, se debe considerar el contexto y características de los estudiantes. Quinto, se deben considerar indicadores de valores agregado cuando sea posible. Finalmente, se deben usar instrumentos y procedimientos estandarizados, válidos, confiables, objetivos y transparentes.

Los indicadores que se utilizan son clasificados en cuatro grupos: (i) resultados SIMCE y estándares de aprendizaje; (ii) medidas de progreso; (iii) otros indicadores de calidad educativa que agrupa variables como autoestima académica de los estudiantes, clima escolar, retención de alumnos, equidad de género, retención, entre otros; (iv) características de los estudiantes que se usan para corregir por nivel socioeconómico un indicador preliminar basado en el resto de los

indicadores. Entre las variables consideradas para este cuarto grupo de indicadores están el nivel de escolaridad de las madres, la vulnerabilidad y grado de aislamiento del establecimiento, la presencia de estudiantes de origen indígena y con necesidades especiales, entre otras.

Para combinar los indicadores la metodología sigue cuatro etapas: (i) construcción inicial de un índice sin corregir por el contexto del establecimiento; (ii) corrección del índice por las características de los estudiantes; (iii) construcción de un índice intermedio que corrige por el contexto socioeconómico de las escuelas; (iv) construcción final del índice de calidad que ajusta la clasificación de algunas escuelas para asegurar que todos los establecimientos entreguen un servicio educativo con estándares mínimos, independientemente de las circunstancias que enfrenta la escuela.

En la primera etapa se combinan los indicadores de resultados de aprendizaje, mejoras y otros indicadores de la calidad. Los estándares de aprendizaje tienen una ponderación del 67% y el resto de los indicadores, en conjunto, tienen un peso de un 33%, distribuido de manera equitativa entre todos ellos. Los estándares de aprendizaje definen lo que los estudiantes deberían saber según los objetivos de aprendizaje estipulados en el currículo vigente para cada materia. Se agregan a través del promedio simple del puntaje estandarizado en cada grado y materia. Luego se definen tres niveles asociados a distintos puntajes estandarizados: (i) Insuficiente; (ii) elemental; (iii) adecuado. Cada uno de estos niveles tiene un puntaje asociado para el cálculo del índice de calidad que corresponde a 0; 65; y 100 respectivamente. Para el cálculo del índice de calidad se consideran los últimos tres años dando una ponderación mayor a los años más recientes.

Del mismo modo, los puntajes de cada prueba SIMCE de los últimos cuatro años se estandariza de manera tal de tener valores entre 0 y 100. Tanto los niveles como las tendencias se usan para el cálculo del índice inicial. En el caso de los niveles, se usan los últimos tres años dando un peso mayor a los años más recientes. En el caso de las tendencias, los últimos cuatro. Las tendencias se calculan para variaciones de un año al siguiente para matemáticas y lectura. Si se determina un aumento estadísticamente significativo en la tendencia, entonces se asigna un valor de 100 para este indicador. Si no hay cambio significativo, un valor de 50 y si hay una caída estadísticamente significativa en los puntajes de SIMCE, un valor de cero.

En el caso de los otros indicadores de la calidad, se consideran ocho indicadores que provienen de dos fuentes de información: datos administrativos del Ministerio de Educación y datos de los cuestionarios contextuales a docentes, estudiantes y padres del SIMCE. Cada uno de los ocho indicadores se puede descomponer en dos a tres dimensiones. En general, cada dimensión está compuesta por la respuesta de estudiantes, docentes o padres a una pregunta con dos a cinco opciones de respuesta (ítems). Si una dimensión particular tiene cinco ítems, éstos se ordenan de manera creciente asignando los valores 0 al ítem uno, 25 al ítem dos, 50 al ítem tres, 75 al ítem cuatro, y 100 al ítem cinco. Una vez transformados los valores de cada ítem, se calcula el valor de la dimensión correspondiente a través del uso de técnicas de análisis factorial. Específicamente, cada ítem tiene una ponderación que depende de su carga factorial y varianza [7, 8]. Adicionalmente, en algunos casos, cada dimensión es el promedio ponderado de los valores correspondientes a estudiantes, padres y profesores.

El proceso descrito anteriormente queda más claro siguiendo el ejemplo del indicador “clima de convivencia escolar” el que está compuesto por tres dimensiones distintas: (i) ambiente de respeto; (ii) ambiente organizado; (iii) ambiente seguro. Cada dimensión se compone de varias preguntas que, en general, tienen 4 ítems. Por ejemplo, para la dimensión ambiente de respeto una pregunta es: “Todos los estudiantes de mi curso nos respetamos,” y sus opciones de respuesta son: “Muy de acuerdo, de acuerdo, en desacuerdo, muy en desacuerdo”. Estas opciones de respuesta se re-escalán para tener valores de 0 a 100. Luego se calcula el ponderador de cada ítem usando técnicas de análisis factorial combinando todas las preguntas e ítems en cada dimensión. Como resultado se tendrá un valor para cada dimensión y cada actor (estudiantes, padres, docentes). En el caso específico del indicador de clima de convivencia escolar, se define una ponderación de 50% para estudiantes, 40% para padres, y 10% para docentes. Por último, para obtener el valor del indicador se calcula el promedio simple de las tres dimensiones.

Una vez que se calculan los valores para cada indicador, éstos se combinan siguiendo los pesos relativos que se definen para cada uno de ellos y se obtiene el indicador inicial que no corrige por el contexto socioeconómico de la escuela. En la segunda etapa, para corregir por el nivel de vulnerabilidad de los estudiantes se estima una regresión lineal multivariada usando datos a nivel de estudiantes. La variable dependiente es el índice sin corregir, las variables de control son todas las características del estudiantes y escuela que controlan por el contexto socioeconómico y el término de error del modelo representa la parte del índice que no depende del contexto. El índice corregido corresponde a la resta entre el índice sin corregir y las características de los alumnos de los establecimientos multiplicadas por los coeficientes estimados en la regresión a nivel individual de estudiantes.

En la etapa tres se definen como límites para cada categoría los percentiles 12; 35; y 85. En otras palabras, los establecimientos que tengan valores del índice final hasta el valor del percentil 12 (inclusive) de la distribución de este índice se clasifican en la categoría de desempeño Insuficiente. Las escuelas que presenten valores del índice final por sobre el valor del percentil 12 y hasta el percentil 35 (inclusive) se clasifican en la categoría de desempeño Medio-Bajo. Los establecimientos que se sitúen por sobre el valor del percentil 35 y hasta el percentil 85 (inclusive) de la distribución del índice final se clasifican en la categoría Desempeño Medio. Finalmente, los colegios con valores del índice final superiores al valor del percentil 85 de la distribución de este índice se clasificarán en la categoría de desempeño alto.

Por último, en la cuarta etapa se definen criterios independientes de las características de los estudiantes para evitar resultados que parecen ser contradictorios. Bajo estos criterios las escuelas son reclasificadas para evitar que existan establecimientos con resultados similares respecto al indicador de estándares de aprendizaje y, al mismo tiempo, en categorías de desempeño alto e insuficiente. También buscan prevenir que escuelas con muy bajos resultados eviten la categoría de desempeño insuficiente sólo porque el contexto social de las familias que atiende es muy desafiante. Con esto se busca promover una focalización de recursos más eficiente pues los establecimientos ordenados en las dos categorías inferiores son priorizados para la entrega de apoyo técnico pedagógico.

Además del apoyo técnico pedagógico que se entrega a las escuelas de desempeño más bajo, la política de aseguramiento de la calidad define sanciones en caso de que los establecimientos no mejoren. Si las escuelas de desempeño insuficiente no muestran una mejora significativa luego de tres años se le entregará información a las familias que contiene datos sobre los treinta establecimientos educacionales más cercanos que estén ordenados en categorías superiores. Para facilitar el acceso a estas escuelas en categorías superiores, se les ofrecerán diversas opciones de transporte. Aún más, si las escuelas no logran mejorar luego de cuatro años, el establecimiento perderá el reconocimiento oficial del Ministerio de Educación y su financiamiento público.

Brasil: Índice de Desenvolvimento de Educação Básica (IDEB)

Desde el 2005, en Brasil se calcula un indicador de calidad educativa a nivel de escuela, municipalidad y estado que se basa en las tasas de aprobación entre primer y octavo grado, y en los resultados de las pruebas estandarizadas de (i) *Prova Brasil* y (ii) *Saeb*, las que son complementarias y en conjunto conforman el *Sistema Nacional de Avaliação da Educação Básica* (SAEB). La *Prova Brasil* se comenzó a usar en 2005. Evalúa las habilidades en lengua portuguesa y matemáticas cada dos años para todas las escuelas públicas ubicadas en zonas urbanas en los grados 4 y 8 que tienen 20 o más estudiantes por clase. *Saeb* es una evaluación muestral que se toma cada dos años desde 1990 y que mide habilidades en las mismas áreas y grados que la *Prova Brasil*, pero agregando el tercer año de enseñanza media. Participan escuelas de las redes pública y privada tanto en zonas urbanas como rurales.

El indicador de calidad que se calcula en Brasil se conoce como el *Índice de Desenvolvimento de Educação Básica* (IDEB) y tiene como objetivo monitorear el rendimiento y nivel de aprendizaje en escuelas y unidades de gestión de la educación escolar tales como municipios y estados. Ha sido diseñado como una herramienta que permite identificar escuelas, municipios y/o estados de bajo desempeño, además de un indicador que sigue la evolución del rendimiento de los estudiantes en escuelas que requieren más apoyo. En términos más aplicados, el indicador intenta medir dos cosas: (i) el nivel de aprendizaje y (ii) si los estudiantes terminan a tiempo la educación básica y no pierden tiempo debido a repitencias o abandonos.

El IDEB toma valores entre 0 y 10, y su cálculo fue definido de manera tal que si una escuela con un valor estandarizado de 5 en sus resultados de la *Prova Brasil* o *Saeb* (según corresponda) tiene estudiantes que demoran dos años en promedio para completar cada grado de educación básica, entonces tendrá un valor de IDEB igual a 5 multiplicado $\frac{1}{2}$, que equivale a 2.5. Asimismo, otra escuela con un valor estandarizado de 5 en la *Prova Brasil* o *Saeb*, con estudiantes que, en promedio, terminan cada grado de educación básica en exactamente un año, entonces su IDEB se mantendrá en 5.

El mecanismo específico de cálculo del IDEB es relativamente simple. En una primera etapa se estandarizan los resultados de aprendizaje de los exámenes de lengua portuguesa y matemáticas de manera tal que tomen valores entre 0 y 10, para luego promediarlos considerando los grados 4 y 8. En una segunda etapa se calcula la tasa de aprobación de cada grado de educación básica

usando los datos del censo escolar. Luego, se suman los inversos de las tasas de aprobación en cada grado y se divide por el número total de grados. Esta medida se interpreta como el tiempo promedio que demoran los estudiantes en terminar un grado. El inverso de este valor está entre 0 y 1 y es el indicador que se usa en el cálculo del IDEB. El valor final del IDEB corresponde a la multiplicación de los valores calculados en ambas etapas.

Uno de los principales usos del IDEB es la identificación de escuelas prioritarias para la asignación de asistencia técnica y financiera a través del *Plano de Desenvolvimento Escolar* o PDE Escola. El gobierno federal de Brasil ofrece una herramienta de planificación a las escuelas de bajo desempeño para apoyarlas en las dimensiones de gestión, relación con la comunidad, infraestructura y definición de su proyecto pedagógico. Estas escuelas usan esta herramienta, elaboran un plan de mejoramiento y reciben recursos para su implementación por dos años. Luego de ese periodo las escuelas pueden seguir usando la herramienta de planificación, pero no tienen financiamiento adicional. Es interesante notar que una evaluación del programa encontró que PDE Escola tiene efectos positivos en el nivel de aprendizaje de los estudiantes, pero sólo cuando las herramientas de planeación están acompañadas de los recursos financieros necesarios para ejecutarlos [9].

El IDEB también sirve como referencia para algunos estados federados de Brasil. Por ejemplo, en Pernambuco se desarrolla una prueba estandarizada que sirve para generar un indicador de calidad por escuela exclusivo del estado y que se calcula siguiendo una metodología equivalente a la del IDEB. Su nombre es *Índice de Desenvolvimento de Educação de Pernambuco* (IDEPE). La ventaja de este indicador en relación con el IDEB es que se basa en datos censales para el tercer año de enseñanza media. De esa forma, el gobierno estadual logra monitorear el desempeño de las escuelas de secundaria en el marco de un sistema de *accountability* que premia a profesores y directores que alcanzan ciertos valores preestablecidos del IDEPE al final de cada año escolar.

Colombia: Índice Sintético de la Calidad Educativa (ISCE)

A fines de 2014 se diseñó el Índice Sintético de la Calidad Educativa (ISCE) con el objetivo de poner a disposición de la comunidad educativa un indicador de calidad de escuelas que fuese fácil de interpretar y que sirviese como insumo para la reflexión y definición de metas específicas a cada establecimiento según el nivel educativo que ofrecen (primaria, secundaria y media). Su elaboración tuvo como principal referencia el IDEB de Brasil. Para la educación básica y secundaria el índice se calcula por medio de la suma ponderada de cuatro indicadores: (i) Progreso de los aprendizajes, que tiene un peso de 40% y valores posibles entre 0 y 4; (ii) desempeño, con un peso de 40% y valores posibles de 0.8 a 4; (iii) eficiencia, medida por la tasa de aprobación. Pesa 10% y toma valores entre 0 y 1; (iv) ambiente escolar, con una ponderación de 10% y valores posibles entre 0 y 1. En el caso de educación media, no se usa el componente de ambiente escolar y el 10% de su ponderación se suma al ponderador asociado al indicador de eficiencia.

El primer indicador de progreso de los aprendizajes se compone de dos partes. La primera tiene una ponderación de 75% y corresponde a la multiplicación de -1 y la tasa de cambio de un año al

siguiente en el porcentaje de estudiantes con un desempeño insuficiente. Por ejemplo, para el caso de primaria, se calcula el promedio entre cuatro valores: las tasas de cambio de la proporción de estudiantes con bajo desempeño en matemáticas y lenguaje, tanto para los grados 3 como 5. Luego este valor se multiplica por -1 para que mayores valores del indicador se asocien a aumentos en la calidad de la escuela. Adicionalmente, se definen un par de criterios para la definición del valor de este indicador: (i) si el colegio presenta un porcentaje de estudiantes con bajo desempeño menor al 10%, se le premia asignándole el máximo puntaje posible; (ii) si el porcentaje de estudiantes con desempeño insuficiente aumenta de un año al siguiente, entonces este componente del indicador de progreso toma el valor de cero.

La segunda parte del componente de progreso tiene un peso relativo del 25% y corresponde al cambio porcentual de un año al siguiente en el porcentaje de estudiantes en el nivel de desempeño avanzado. Al igual que para el caso del cambio en el porcentaje de estudiantes de bajo desempeño, se definen dos criterios adicionales. Si en la escuela más de un 90% de los estudiantes tiene un nivel avanzado entonces se le asigna el valor máximo posible. Por el contrario, si este porcentaje disminuye de un año al siguiente, el indicador toma el valor de cero.

El componente de desempeño se mide por el resultado de matemáticas y lenguaje en la prueba SABER.³ En primaria, el promedio de los resultados para tercer y quinto grado en ambas asignaturas entrega el valor final que contribuirá al indicador de calidad final. La escala de cada una de estas pruebas va desde 100 a 500, por lo que este valor se re-escala multiplicándolo por 4/500, lo que da un rango entre 0.8 y 4.

La tercera dimensión de eficiencia equivale a la tasa de aprobación en cada nivel educativo. Tiene un valor entre 0 y 1 e incluye ajustes realizados por el Ministerio de Educación para corregir por los estudiantes que se retiran o desertan del sistema en medio del año escolar. Esto es posible gracias al sistema integrado de matrícula que registra y reporta mensualmente la cantidad de estudiantes en el sistema escolar.

Finalmente, el componente de ambiente escolar se calcula a partir de dos medidas que provienen del cuestionario de acciones y actitudes ciudadanas de la parte no cognitiva de la prueba de competencias ciudadanas que toman los estudiantes de quinto y noveno grado cuando participan en la evaluación de aprendizajes SABER: (i) ambiente en el aula, y (ii) seguimiento al aprendizaje. Cada medida tiene una escala que va de 100 a 200 para cada uno de los grados 5 y 9. Las cuatro mediciones resultantes son promediadas y ajustadas de manera tal de obtener una escala con valores entre 0 a 1.

El ISCE se calcula para las escuelas que tienen resultados de la prueba SABER en los grados tercero, quinto y noveno/onceavo. Si un colegio tiene datos sólo para matemáticas o lenguaje, los resultados de la asignatura faltante son imputados a partir de la información disponible de la otra asignatura. Cabe destacar que, a pesar de que la prueba SABER sólo reporta oficialmente el

³ Las pruebas SABER forman parte de la evaluación anual que realiza el Ministerio de Educación de Colombia a los estudiantes en los grados tercero, quinto, noveno y onceavo. Mide los conocimientos en matemáticas y lenguaje en los cuatros grados, y se agregan los conocimientos en ciencias naturales y competencias ciudadanas en los grados quinto, noveno y onceavo. Por último, se añade una prueba de inglés en el caso del grado onceavo.

promedio de los puntajes a nivel de escuela si hay más de seis estudiantes que rinden los exámenes, para fines del cálculo de ISCE sí se consideran los resultados en escuelas con menos estudiantes.

El puntaje obtenido en el ISCE tiene como consecuencia la definición de una meta de mejoramiento mínimo anual específica a cada establecimiento. De esta forma se busca incentivar que cada colegio busque superarse a sí mismo y aumente sus resultados de aprendizaje de forma individual, pero siempre teniendo en mente el logro de la meta que este país se ha puesto para su sistema educativo: ser el país mejor educado de América Latina en el año 2025.

Un breve resumen de los casos de América Latina

Los índices revisados de Chile, Brasil y Colombia tienen distintos objetivos y usan diversas variables y metodologías para combinar la información disponible sobre calidad escolar. La Tabla 5 resume los tipos de indicadores que se usan y la ponderación que se define para cada dimensión. Al igual que en los sistemas fuera de América Latina, los indicadores más importantes son los relacionados con las pruebas estandarizadas de aprendizaje. Específicamente, entre un 50 a 80% del peso de los indicadores de calidad de escuelas corresponde al nivel y mejoras de aprendizaje medidas por estas evaluaciones.

Tabla 5: Pesos asociados a las variables consideradas en el cálculo de indicadores de calidad en América Latina

Elemento	Chile			Brasil - IDEB	Colombia - ISCE
	SNED ^(a)	SEP ^(b)	SAC		
Nivel de aprendizaje	37%	70%	70.3%	50%	40%
Tasa de aprobación y retención	7%	15%	3.3%	50%	10%
Mejoras de aprendizaje	28%	-	3.3%	-	40%
Medidas de Equidad	15%	-	3.3%	-	-
Ambiente escolar	-	-	3.3%	-	10%
Otros	13%	15%	16.5%	-	-

(a): Los indicadores se comparan entre escuelas similares en cuanto a su ubicación, nivel de enseñanza y nivel socioeconómico. Estas variables están incluidas implícitamente en la construcción de 125 grupos de escuelas comparables

(b): Se reportan las ponderaciones asociadas al Índice de Calidad Educativa usado en la clasificación de escuelas "En Recuperación"

4. Reflexiones para el diseño y uso de índices de calidad escolar

Varios gobiernos en ALC se enfrentan a la desafiante tarea de dar un uso significativo a la extensa y diversa información sobre resultados escolares que han venido recopilando durante los últimos años. La generación de un índice de calidad escolar tiene el potencial de convertirse en una herramienta efectiva para identificar las escuelas de mejor y peor desempeño sostenido en el tiempo, focalizar recursos de manera más eficiente, y entregar información relevante a la gestión y dirección en cada escuela para una mejor toma de decisiones. Para lograr estos objetivos es clave

diseñar índices informativos que tengan una interpretación simple que logre guiar de forma efectiva las prácticas de las escuelas y actores principales a nivel escolar. Del mismo modo, se requiere la definición de usos para el índice que sean validados por la comunidad escolar, aumente el compromiso con el aprendizaje de estudiantes y que promueva acciones valoradas por las escuelas, directivos y profesores.

Diseño de índices de calidad

La revisión de la experiencia internacional es un ejercicio útil para identificar las mejores prácticas en la construcción y diseño de índices informativos, válidos y aceptados por la comunidad educativa. Nuestra revisión sugiere que existen varios factores relevantes a considerar:

- i. La concepción de calidad de escuelas debería incorporar múltiples dimensiones tales como resultados de aprendizaje, retención, desarrollo de habilidades sociales, elementos de la cultura educativa e incentivos para que las escuelas implementen sus propios sistemas de medición y evaluación de procesos educativos;
- ii. La elección de los indicadores y la valoración que se le da a cada dimensión difiere en cada sistema y también entre primaria y secundaria dentro de cada sistema. Estas diferencias, parecen reflejar de forma implícita una combinación entre la concepción de calidad que tienen las autoridades sobre el servicio de las escuelas y la calidad de los sistemas de información con que se cuentan. En la mayoría de los casos revisados la ponderación asignada a las pruebas estandarizadas es mayor en primaria que en secundaria. En este último nivel, la importancia relativa de la permanencia en el sistema escolar aumenta a costa de la importancia relativa de los indicadores de aprendizaje.
- iii. Se debe tomar en cuenta factores técnicos en la construcción de indicadores de calidad de escuelas, especialmente, poder asegurar una alta precisión estadística de los indicadores usados. Los resultados de pruebas estandarizadas pueden ser poco informativos en el caso de escuelas con pocos estudiantes. Una investigación que usó datos de Carolina del Norte entre 1994 y 2000 muestra que las escuelas más pequeñas del sistema eran 23 veces más probables de pertenecer al 25% de escuelas que más mejora sus resultados en pruebas estandarizadas de un año al siguiente cuando se comparan con las escuelas más grandes [10]. Los autores del estudio argumentan que esta diferencia se debe a que los resultados de aprendizaje en escuelas pequeñas son medidas de la calidad mucho menos confiables. En particular, sostienen que factores transitorios, los cuales no representan la capacidad de la escuela para mejorar aprendizajes pero que sí afectan los puntajes de las pruebas estandarizadas, son más relevantes en escuelas con menos estudiantes. Por ejemplo, la presencia de un estudiante excepcionalmente bueno o malo para tomar pruebas estandarizadas en un año particular, o que haya ruidos molestos en la sala cuando se toma la prueba, pueden afectar mucho más a los puntajes en escuelas pequeñas que en escuelas más grandes.

Otras investigaciones similares basadas en países de ALC muestran patrones similares: usar un número bajo de observaciones para medir la calidad de las escuelas puede ser poco

confiable, pues podrían estar reflejando situaciones fortuitas o fuera del control de la escuela. En términos más técnicos, uno de los estudios basado en datos de inicios de la década de 1990s en Chile, encuentra que, para una evaluación a 15 estudiantes, los factores transitorios representan un 30 y 44% de la variabilidad de los resultados de matemáticas y lenguaje, respectivamente. Estos porcentajes caen a 33 y 21% cuando el número de estudiantes evaluados aumenta a 30 [11]. En general la evidencia empírica disponible sugiere que combinar resultados de aprendizaje de varios años logra reducir considerablemente este componente transitorio y refleja mejor el desempeño real de la escuela.

- iv. Se debería considerar la posibilidad de introducir medidas de valor agregado que midan la trayectoria del aprendizaje de cada estudiante por varios años. De esta forma, el indicador es más efectivo para informar sobre la capacidad de las escuelas para incrementar las habilidades iniciales de los estudiantes y, al mismo tiempo, reduce el efecto de factores contextuales como el nivel socioeconómico de las familias. En esta línea, existen estimaciones basadas en escuelas de Holanda que muestran que más de un 35% de la variabilidad de los resultados de aprendizaje a nivel de escuela para un grado en particular, se explica por niveles de aprendizajes de sus estudiantes en años anteriores [12]. Esto sugiere que basar índices en niveles más que progreso podría entregar información distorsionada de la calidad de la escuela, entendida como su capacidad para mejorar el nivel inicial de conocimientos y habilidades de los estudiantes [13].
- v. Deberían incluirse variables que miden diversas dimensiones de la calidad del servicio que entregan las escuelas para así reflejar el trabajo de los establecimientos educativos de una forma más integral y no centrarse en exceso en resultados de pruebas estandarizadas. No obstante, se debe tener presente que tanto las variables elegidas como la forma en que se combinen deben ser lo suficientemente simples para que cumplan con el requisito de crear indicadores informativos y comprensibles para directivos y docentes. Si se incluyen múltiples medidas, se deberían entregar indicaciones claras y transparentes sobre cómo la escuela podría mejorar cada dimensión. De esta manera, es más probable inducir cambios concretos por parte de los actores principales del sistema escolar.
- vi. Acompañar el índice con indicaciones claras y transparentes sobre cómo las escuelas pueden mejorar en cada dimensión medida. En este sentido, podría ser efectivo seguir estrategias que involucren a los mismos profesionales de la educación en la recolección, uso y evaluación de datos relacionados a la calidad educativa. Además de los casos de Hong Kong y Holanda, otros sistemas también han comenzado a incursionar en esta línea. Por ejemplo, la iniciativa “Aprendizaje visible” implementada en algunas escuelas en Wisconsin, Estados Unidos, consiste en promover una cultura de autoevaluación y aprendizaje continuo a través de la reflexión y análisis de las propias prácticas [14]. Específicamente, en este programa los profesores graban sus propias clases, identifican sus fortalezas y debilidades, las dinámicas de la clase y el grado de compromiso de los estudiantes, para luego discutir con un asesor estrategias para mejorar sus prácticas. Aquellos sistemas que incluyen medidas relacionadas a la evaluación docente en sus indicadores de calidad escolar podrían considerar autoevaluaciones basadas en el análisis

de videos que docentes podrían hacer de sus propias clases y/o las evaluaciones de estos videos realizadas por asesores o supervisores.

- vii. La discusión sobre las consecuencias asociadas a diversos valores del indicador debiera ser inclusiva y considerar la visión de distintos actores. Es clave asegurar que los esfuerzos realizados por las escuelas se vean reflejados en los cambios del indicador que mide su calidad, evitando establecer castigos que no están vinculados a las acciones de las escuelas o maestros.

Uso de índices de calidad

La definición de las consecuencias asociadas a diversos valores del indicador es otro punto clave. Existe un intenso debate sobre cuál es la intervención más efectiva para mejorar el nivel de aprendizajes, especialmente en el caso de las escuelas de bajo desempeño persistente. Encontrar el balance adecuado entre entregar más apoyo y establecer una mayor exigencia es un desafío complejo para las autoridades. La revisión de la experiencia internacional muestra que las consecuencias usualmente definidas para escuelas de bajo desempeño persistente se pueden dividir en dos grupos principales.

En el primer grupo se encuentran medidas menos disruptivas que no obligan la reorganización del establecimiento educativo. Se enfocan principalmente en proveer incentivos, apoyo y/o herramientas para que las escuelas puedan implementar planes de mejoramiento o reestructuren sus procesos administrativos y pedagógicos de manera voluntaria. Entre estas medidas están:

- (i) Entrega de información a familias para que aumenten su monitoreo de las actividades de las escuelas de sus hijos;
- (ii) implementación de planes de mejoramiento;
- (iii) asistencia técnica que puede ser provista por parte del estado o privados;
- (iv) reorganización voluntaria del establecimiento que puede considerar la reforma de elementos claves como el currículo, gestión o docencia.

Dentro del segundo grupo se encuentran medidas más disruptivas que se caracterizan por una reestructuración obligatoria de las prácticas de la escuela, o incluso su cierre definitivo. Entre estas se cuentan:

- (i) Cambio del director;
- (ii) ofrecer a familias la transferencia a otras escuelas de mayor desempeño;
- (iii) reconstitución de la escuela, la que implica la remoción de gran parte del personal, además de implementar un nuevo proyecto educativo;
- (iv) fusión de escuelas de bajo y alto rendimiento que comienzan a compartir un mismo cuerpo directivo;
- (v) toma de control, en donde la dirección y administración del establecimiento pasa a una agencia especializada del estado o de un privado;
- (vi) sanciones financieras como en el caso de Holanda; y

- (vii) La más fuerte medida es el cierre definitivo del establecimiento y reasignación de los estudiantes a escuelas de mejor rendimiento promedio.

Una de las consecuencias que se adoptan más comúnmente es la elaboración e implementación de planes de mejoramiento. Un caso interesante es el de Inglaterra. En su sistema escolar se diseñó un programa a cargo de la *Office of Standards in Education, Children's Services and Skills* (OFSTED) que tiene como objetivo incrementar el rendimiento de las escuelas y su calidad a partir de inspecciones regulares. Para lograrlo, evaluadores externos realizan visitas periódicas a establecimientos para observar y recopilar información cualitativa de los procesos organizativos y pedagógicos de las escuelas. Un reporte identifica las principales debilidades de cada establecimiento y lo entrega a las escuelas, quienes deben tomar en cuenta las recomendaciones y elaborar planes de mejoramiento que sean implementados por los siguientes cuatro años. El plan debe ser elaborado por la dirección de cada escuela y debe incluir una planificación detallada de las acciones que tomarán en respuesta al diagnóstico de los evaluadores de la OFSTED estableciendo plazos y metas claras que serán evaluadas en la siguiente visita. Si en la siguiente visita, la OFSTED considera que la escuela no ha logrado una mejora significativa, la escuela enfrentará consecuencias más disruptivas.

Al momento de definir medidas más disruptivas, un punto importante para considerar es que las consecuencias que se elijan deberían ser validadas por la comunidad escolar. En este sentido, la experiencia chilena durante el año 2010 es ilustrativa. En ese año, el Ministerio de Educación intentó involucrar más a los padres a través de la entrega de una carta informativa que mostraba los resultados de aprendizaje promedio en su escuela de una manera bastante simple. Siguiendo los mismos colores de un semáforo, las escuelas por sobre el promedio nacional en las pruebas estandarizadas de matemáticas y lenguaje, se pintaban de verde, las cercanas al promedio de amarillo, y las que estaban bajo el promedio de rojo. La carta informativa era acompañada de un mapa que mostraba todas las escuelas del barrio pintadas con alguno de estos tres colores.

El indicador de calidad escolar usado no tomó en consideración la base de conocimientos con que los estudiantes ingresan a la escuela ni tampoco información sobre el contexto social de las familias atendidas. Esto generó que la medición de calidad estuviera altamente correlacionada con el nivel socioeconómico del barrio donde se ubicaba la escuela, lo que provocó una reacción adversa por la mayor parte de la comunidad escolar, que la consideró injusta por no reflejar el nivel de esfuerzo de profesores ni directivos y por contribuir a estigmatizar los establecimientos más vulnerables, especialmente los públicos. Esta visión de la política de los semáforos dominó la opinión pública, y forzó a su revocación en menos de un año desde su implementación.

Experiencias como la de los semáforos de la educación en Chile no sólo reflejan la necesidad de diseñar índices precisos e informativos, sino que también la de crear un sistema de medición y evaluación de la calidad escolar que conecte indicadores a consecuencias validadas por la comunidad escolar y que sean percibidas como justas en el sentido que no se establezcan castigos por resultados que están fuera del control de las escuelas.

Referencias

- [1] M. Ehren y N. Shackleton, «Risk-based school inspections: impact of targeted inspection approaches on Dutch secondary schools,» *Educational Assessment, Evaluation and Accountability*, vol. 28, pp. 299-321, 2016.
- [2] Education Inspectorate, «Inspection Framework Primary Education,» Ministry of Education, Culture and Science, Amsterdam, 2018.
- [3] X. Ni, A. Bowers y J. Esswein, «What Counts in Calculating School and District Level Performance Index Scores: A Summary and Analysis of Academic Performance Index Metrics across the 50 States,» Teachers College, Columbia University, New York, 2016.
- [4] C. Martin, S. Sargrad y S. Batel, «Making the grade: A 50-State Analysis of School Accountability Systems,» Center for American Progress, Washington, D.C., 2016.
- [5] E. Davidson, R. Reback, J. Rockoff y H. L. Schwartz, «Fifty ways to leave a child behind: Idiosyncrasies and discrepancies in states' implementation of NCLB,» *Educational Researcher*, , vol. 44, n° 6, pp. 347-358, 2015.
- [6] G. Elacqua, M. Martínez, H. Santos y D. Urbina, «Short-run effects of accountability pressures on teacher policies and practices in the voucher system in Santiago, Chile,» *School Effectiveness and School Improvement*, vol. 27, n° 3, pp. 385-405, 2016.
- [7] Agencia de la Calidad de la Educación, Chile, «Metodología de Construcción: Indicadores de desarrollo personal y social,» Agencia de la Calidad de la Educación, Santiago, 2016.
- [8] T. Brown, *Confirmatory Factor Analysis for Applied Research*, New York: The Guilford Press, 2015.
- [9] F. Alves, G. Elacqua, M. Martínez y H. Santos, «Efeitos do Plano de Desenvolvimento da Escola nos resultados escolares,» *Estudos em Avaliação Educacional*, vol. 27, n° 64, pp. 128-159, 2016.
- [10] T. J. Kane y D. O. Staiger, «The promise and pitfalls of using imprecise school accountability measures.,» *Journal of Economic perspectives*, vol. 16, n° 4, pp. 91-114, 2002.
- [11] K. Y. Chay, P. J. McEwan y M. Urquiola, «The central role of noise in evaluating interventions that use test scores to rank schools,» *American Economic Review*, vol. 95, n° 4, pp. 1237-1258, 2005.

- [12] A. C. Timmermans, S. Doolaard y I. De Wolf, «Conceptual and empirical differences among various value-added models for accountability,» *School Effectiveness and School Improvement*, vol. 22, n° 4, pp. 393-413, 2011.
- [13] E. San Martín y A. Carrasco, «Clasificación de escuelas en la nueva institucionalidad: contribución de modelos de valor agregado para una responsabilización justa,» Centro de Políticas Públicas UC, Santiago, 2012.
- [14] Edutopia, «Using Video for Professional Development,» George Lucas Educational Foundation, 2018. [En línea]. Available: <https://www.edutopia.org/video/using-video-professional-development>. [Último acceso: 01 Noviembre 2018].