

Daniel Pinkasz



unesco

Instituto Internacional
de Planeamiento de
la Educación

El uso de los resultados de las pruebas de evaluación de los aprendizajes en el planeamiento de las políticas educativas en seis países de la región

Informe regional



Daniel Pinkasz

El uso de los resultados de las pruebas de evaluación de los aprendizajes en el planeamiento de las políticas educativas en seis países de la región

Informe regional

Advertencia

El uso de un lenguaje no discriminador en función del género de las personas es una de las prioridades de la Oficina para América Latina del Instituto Internacional de Planeamiento de la Educación (IIEP) de la UNESCO. Actualmente, su aplicación en la lengua española plantea soluciones muy distintas, sobre las que aún no se ha logrado ningún acuerdo. En tal sentido, evitamos usar en nuestros textos expresiones tradicionales que ya han sido abandonadas en el español académico y profesional contemporáneo, tales como la palabra “hombres” para referirse a un conjunto de seres humanos. Además, siempre que es posible, procuramos emplear palabras de género neutro en reemplazo de palabras en género masculino. Sin embargo, con el fin de evitar la sobrecarga gráfica que supondría utilizar desdoblamientos léxicos en artículos, sustantivos y adjetivos para subrayar la existencia de distintos géneros, hemos optado por el uso genérico del masculino, en el entendimiento de que todas sus menciones representan siempre a todas las personas.

.....

Esta publicación está disponible en acceso abierto bajo la licencia Attribution-ShareAlike 3.0 IGO (CC- BY- SA 3.0 IGO) (<http://creativecommons.org/licenses/by-sa/3.0/igo/>). Al utilizar el contenido de la presente publicación, los usuarios aceptan las condiciones de utilización del Repositorio UNESCO de acceso abierto (www.unesco.org/open-access/terms-use-ccbysa-sp).

.....

Los términos empleados en esta publicación y la presentación de los datos que en ella aparecen no implican toma de posición alguna de parte de la UNESCO o del IIEP en cuanto al estatuto jurídico de los países, territorios, ciudades o regiones ni respecto de sus autoridades, fronteras o límites.

.....

Publicado en 2021 por la Oficina para América Latina del Instituto Internacional de Planeamiento de la Educación de la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura.

Agüero 2071, Ciudad Autónoma de Buenos Aires, Argentina.

© UNESCO 2021



unesco

Instituto Internacional
de Planeamiento de
la Educación

Agradecimientos

El presente estudio se desarrolló bajo la coordinación general del área de Investigación y Desarrollo de la Oficina para América Latina del IIEP UNESCO, a cargo de Fernanda Luna (coordinadora interina), Felicitas Acosta y Bianca Gentinetta Delfino. Se agradece a Néstor López, por su trabajo en los inicios del proyecto y sus aportes en la lectura y comentarios al mismo. También se agradece el apoyo del IIEP UNESCO París, especialmente a leva Raudonyte por los intercambios y contribuciones al contenido del proyecto.

La redacción de este informe fue desarrollada por Daniel Pinkasz (investigador principal) con el apoyo de un equipo de consultores a cargo de los estudios de caso que conforman la investigación: Nancy Montes, Alejandra Meraz Velasco, Jorge E. Castillo Peña, Cinthia Chiriboga, Mariana E. Castro y Margarita Romero García.

A todos y todas, nuestro mayor agradecimiento por el compromiso, la dedicación y el profesionalismo en el trabajo realizado.

Contenidos

Listado de siglas	6
Resumen Ejecutivo	9
1. Presentación del estudio	22
Objetivos del proyecto	24
2. Enfoque teórico y metodológico	25
Las relaciones entre políticas públicas y planificación en América Latina, algunas reflexiones para determinar su encuadre	25
Estrategia metodológica	29
Acerca de este informe	31
3. Trayectorias comparadas de los dispositivos de evaluación de aprendizajes a gran escala abarcadas por el estudio.	
Un marco de interpretación	33
Evaluaciones estandarizadas: dos funciones en tensión	33
Una interpretación de la trayectoria de los dispositivos de evaluación en los seis países estudiados	37
4. Usos previstos de las evaluaciones estandarizadas según los marcos normativos	46
Funciones asignadas por los encuadres normativos	46
Encuadres normativos y usos previstos en los países del estudio	47
A modo de síntesis	60
5. Usos efectivos de las evaluaciones a gran escala	63
Usos efectivos reportados de los resultados	64
5.1. Planificar estrategias	64
5.2. Producir conocimiento (identificar problemas educativos, áreas de intervención y asociar factores con los resultados de los aprendizajes)	68

5.3. Desarrollar políticas y programas de apoyo, generales y focalizados, para la mejora de los resultados	71
5.4. Orientar la definición del currículo y la elaboración de materiales pedagógicos	79
5.5. Formar y fortalecer la profesionalización de los docentes	81
5.6. Certificar estudios y realizar pruebas de selección para el ingreso a la educación superior	83
5.7. Evaluar y/o incentivar a docentes	84
5.8. Ordenar posiciones según niveles de desempeño	85
Usos predominantes	86
6. Factores que explican el uso	94
Factores institucionales	95
Factores técnicos	104
Factores políticos	108
Factores comunicacionales	114
7. Conclusiones	122
8. Recomendaciones	127
Anclar los usos de los resultados en acuerdos sobre las funciones de los sistemas de evaluación	127
Proyectar de manera progresiva e incremental los usos de las evaluaciones en el mediano plazo	128
Mantener continuidad en las mediciones	129
Alinear usos con las características técnicas de las pruebas	129
Definir políticas de uso de los resultados	130
Distinguir entre audiencias y usuarios	130
Fortalecer las capacidades técnicas de los estados subnacionales y de los usuarios locales	131
Bibliografía	132
Anexos	135
Anexo 1: Criterios de selección de actores entrevistados	135
Anexo 2: Características de las evaluaciones por país	138

Listado de siglas

ANEP	Administración Nacional de Educación Pública (Uruguay)
ANMEB	Acuerdo Nacional para la Modernización de la Educación Básica (México)
APIT	Análisis de Presencia Institucional en el Territorio (Ecuador)
BNCC	Base Nacional Comum Curricular (Brasil)
CABA	Ciudad Autónoma de Buenos Aires (Argentina)
CFE	Consejo Federal Educativo (Argentina)
CINE	Clasificación Internacional Normalizada de la Educación
CODICEN	Consejo Directivo Central (Uruguay)
CONSED	Conselho Nacional de Secretários de Educação
DIEE	División de Investigación, Evaluación y Estadística (Uruguay)
ERCE	Estudio Regional Comparativo Explicativo
FEPBA	Finalización de Estudios Primarios de la Ciudad de Buenos Aires (Argentina)
FESBA	Finalización de Estudios Secundarios de la Ciudad de Buenos Aires (Argentina)
FNDE	Fondo Nacional de Desarrollo de la Educación (Brasil)
FUNDAE	Fundo para o Desenvolvimento e Autonomia da Escola (Sobral-CE, Brasil)
FUNDECREI	Fundo para o Desenvolvimento e Autonomia dos Centros de Referência de Educação Infantil (Brasil)
FUNDEB	Fundo de Manutenção e Desenvolvimento da Educação Básica e de Valorização dos Profissionais da Educação (Brasil)
ICCS o CIVIC o CÍVICA	Estudio Internacional de Civismo y Ciudadanía (International Civic and Citizenship Education Study)
IDE	Índice de Desempeño Escolar

IDEB	Índice de Desenvolvimento da Educação Básica (Brasil)
INEE	Instituto Nacional para la Evaluación de la Educación de México
INEEd	Instituto Nacional de Evaluación Educativa de Uruguay
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Brasil)
INEVAL	Instituto Nacional de Evaluación de Ecuador
ISEP	Instituto Superior de Estudios Pedagógicos (Córdoba, Argentina)
LEN	Ley de Educacional Nacional (Argentina)
LLECE	Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación
LOEI	Ley Orgánica de Educación Intercultural (Ecuador)
MEJOREDU	Comisión Nacional para la Mejora Continua de la Educación (México)
ONE	Operativo Nacional de Evaluación (Argentina)
PAR	Plano de Ações Articuladas (Brasil)
PEEME	Programas Estatales de Evaluación y Mejora Educativa (México)
PISA	Programa para la Evaluación Internacional de Alumnos (OCDE)
PNEE	Política Nacional de Evaluación Educativa (México)
POAI	Plan Operativo Anual Integral (Argentina)
PDDE	Programa Dinheiro Direto na Escola (Brasil)
SAC	Sistema de Aseguramiento de la Calidad (Chile)
SAEB	Sistema de Avaliação da Educação Básica (Brasil)
SARESP	Sistema de Avaliação de Rendimento Escolar do Estado de São Paulo (Brasil)
SEIE	Secretaría de Evaluación e Información Educativa (Argentina)
SENPLADES	Secretaría Nacional de Planificación y Desarrollo (Ecuador)
SEP	Secretaría de Educación Pública (México)
SER	Sistema de Evaluación y Rendición Social de Cuentas (Ecuador)

SERSE	Segundo Estudio Regional Comparativo Explicativo
SIMCE	Sistema de Medición de la Calidad Educativa (Chile)
SNED	Servicio Nacional de Evaluación de Desempeño (Chile)
SNEE	Sistema Nacional de Evaluación Educativa (México)
SPAECE	Sistema Permanente de Avaliação da Educação Básica do Ceará (Brasil)
TERCE	Tercer Estudio Regional Comparativo Explicativo
TESBA	Finalización de Estudios Terciarios de la Ciudad de Buenos Aires (Argentina)
UNDIME	União Nacional dos Dirigentes Municipais de Educação (Brasil)

Resumen Ejecutivo

Presentación

Este informe presenta los principales resultados de un estudio comparado realizado sobre los usos de los datos proporcionados por las pruebas de evaluación de los aprendizajes a gran escala de Argentina, Brasil, Chile, Ecuador, México y Uruguay. Asimismo, propone una interpretación de los resultados de los seis estudios nacionales a cuya lectura se remite.

El propósito de este trabajo es generar conocimiento específico y contextualizado para orientar el uso de resultados de evaluación de aprendizajes de gran escala de manera constructiva en el planeamiento estratégico y en la gestión de los sistemas educativos en distintos países de América Latina.

A partir de la iniciativa del estudio Use of Learning Assessment Data in the Planning Cycle realizado por la sede del IIEP UNESCO, en París, que pone el foco en los factores político-económicos que influyen en el uso de resultados de evaluaciones de gran escala en la región subsahariana de África, se planteó la conveniencia de desarrollar un estudio semejante en países de Latinoamérica con el doble propósito de generar un conocimiento contextualizado para los países de la región y de establecer comparaciones e intercambios de experiencias entre ambas regiones.

Objetivo general

Indagar sobre la relación entre los usos de los datos de los operativos de evaluación de aprendizaje previstos en los marcos legales y en el diseño de los dispositivos, junto con su justificación y el uso efectivo que se hace de esos datos, con el fin de generar conocimiento que permita un mayor

aprovechamiento del potencial de estos instrumentos en el diseño de las políticas educativas de la región.

Objetivos específicos

- Indagar sobre el uso de los resultados de los operativos de evaluación de aprendizajes comprometido por los organismos de planificación y gestión de la política educativa de cada país.
- Identificar el uso efectivo que estos organismos hacen de los resultados de las evaluaciones en distintas fases del planeamiento de las políticas educativas.
- Identificar qué factores influyen en los usos efectivos de los resultados de la evaluación para el planeamiento de políticas.

Antecedentes

Se realizó un estudio de casos orientado a la comprensión de los procesos que subyacen al uso observado de los resultados de las evaluaciones en distintos niveles del sistema educativo. En efecto, los seis países indagados proporcionan una variedad de características que resultan apropiadas para la comparación. Tres Estados son federales, Argentina, Brasil y México. Los seis países exhiben variadas trayectorias de institucionalización de sus dispositivos de evaluación y han definido diferentes modelos evaluativos desde el punto de vista de las consecuencias que originan los resultados de las evaluaciones estandarizadas. En cada país, se seleccionaron las evaluaciones nacionales estandarizadas de logro de aprendizajes, aplicadas en los niveles de educación obligatoria (CINE 1, 2 y 3) y las principales evaluaciones internacionales, es decir, las aplicadas por el LLECE, PISA y otras (según el país).

Uso de datos de las evaluaciones a gran escala

Si bien cada país tiene predominio y modalidades específicas de empleo de datos, en su conjunto los usos reportados pueden encuadrarse en las siguientes categorías.

1. Planificar estrategias en distintos niveles del sistema.
2. Producir conocimiento (identificar problemas educativos, áreas de intervención y asociar factores con los resultados de los aprendizajes).
3. Desarrollar políticas y programas de apoyo, generales y focalizados, para la mejora de los resultados.
4. Orientar la definición del currículo y la elaboración de materiales pedagógicos.
5. Formar y fortalecer la profesionalización de los docentes.
6. Certificar estudios y realizar pruebas de selección para el ingreso a la educación superior.
7. Evaluar y/o incentivar a docentes.
8. Ordenar posiciones según niveles de desempeño.

A continuación, se presenta una tabla con el empleo de datos por país.

Usos efectivos	País
Planificar estrategias en distintos niveles del sistema.	TODOS
Producir conocimiento (identificar problemas educativos, áreas de intervención y asociar factores con los resultados de los aprendizajes).	TODOS
Desarrollar políticas y programas de apoyo, generales y focalizados, para la mejora de los resultados.	TODOS
Orientar la definición del currículo y la elaboración de materiales pedagógicos.	TODOS
Formar y fortalecer la profesionalización de los docentes.	TODOS

Certificar estudios y realizar pruebas de selección para el ingreso a la educación superior.	Brasil, Ecuador, México
Evaluar y/o incentivar a docentes.	Chile, Brasil, México y entidades subnacionales de Brasil y México
Ordenar posiciones (escuelas o países) según niveles de desempeño.	Para Evaluaciones Nacionales: Chile, Brasil, México, entidades subnacionales de Brasil y México. Para evaluaciones internacionales: TODOS

Fuente: elaboración propia sobre la base de los Informes Nacionales.

Principales hallazgos y recomendaciones

Factores institucionales

En Argentina, la evaluación muestra una trayectoria de consolidación en cuanto a su permanencia y también de complejización técnica, asociada con un federalismo de concertación a través de su Consejo Federal de Educación. Los vaivenes de la trayectoria de las evaluaciones que registra este país parecieran estar ligados a la discusión acerca de las funciones de la evaluación y la noción de calidad de la educación. Las controversias en Argentina pasan por la construcción de un modelo de evaluación que asuma una noción amplia o una noción estrecha de la calidad de la educación, según los períodos, en tanto parece que la no resolución de esta controversia a nivel institucional posterga la definición del empleo de los resultados en la macropolítica, lo que no impide el despliegue de usos específicos en distintos niveles del sistema y en las provincias.

En Brasil, el Instituto Nacional de Estudios e Investigaciones Educativas (INEP) se destaca por su trayectoria de institucionalización y por la densidad y coordinación de su sistema de relaciones institucionales, el que se encuentra principalmente al servicio de la coordinación de la federación con las entidades subnacionales.

En el caso de Chile, la Agencia de la Calidad de la Educación fue creada cuando el sistema de evaluación presentaba una extensa y consolidada trayectoria. El Ministerio de educación y la Agencia exhiben un alto nivel de coordinación. Por lo tanto, la división de tareas entre ambos organismos ha posibilitado una mayor complejidad en el análisis de los resultados y en la lectura, tanto para el nivel micro como para el meso. Al mismo tiempo, se observa que habría disminuido la capacidad ministerial para actuar a partir de esos resultados. Las relaciones entre la Agencia y las unidades usuarias del Ministerio de Educación se desarrollan con fluidez, si bien se reporta cierto solapamiento de coordinación en la función de acompañamiento a las escuelas, donde se producen superposiciones de roles.

En Ecuador, el Instituto Nacional de Evaluación (INEVAL) presenta una trayectoria relativamente reciente y los ajustes que se reportan parecieran responder a búsquedas de equilibrio propias de una fase de instalación. Así se advierte un impulso inicial que pareciera haber menguado con el cambio de autoridades educativas nacionales en 2017 y la ausencia de directrices estratégicas sobre las cuales afirmar las funciones iniciales de las pruebas de evaluación. También se observaron algunas tensiones menores en la coordinación entre el Instituto y el Ministerio de Educación. Desde el ministerio se reclama una mayor participación en la elaboración de las pruebas, a lo que las autoridades del INEVAL responden con la afirmación de su autonomía. Otra limitación institucional que se señala para el uso de las evaluaciones es la ausencia de focos relevantes en la evaluación, como el de la educación intercultural bilingüe, y una apertura por grupos de poblaciones prioritarias.

En México, las tensiones generadas por el uso de los resultados en pos de implementar un sistema de premios y castigos para la carrera docente y por las atribuciones asignadas al Instituto Nacional para la Evaluación de la Educación (INEE) para proporcionar directivas a las entidades federativas,

habrían generado problemas de coordinación y minado la legitimidad técnica del Instituto frente a las entidades subnacionales y frente a las representaciones docentes. A la vez, podrían haber puesto en duda sus atribuciones institucionales y su autonomía como organismo, hasta que, finalmente, la tensión desembocó en su cierre.

En Uruguay, existen dos organismos que comparten las responsabilidades sobre las evaluaciones: el Instituto Nacional de Evaluación Educativa (INEEd), un organismo autónomo, y la División de Investigación, Evaluación y Estadística (DIEE). Distintas fuentes entrevistadas coinciden en señalar una superposición de sus funciones. No obstante, este solapamiento no impide el desarrollo de una política sostenida de evaluación. La convivencia de ambas unidades puede leerse como la manifestación institucional de la coexistencia de diferentes posiciones en relación con la opción entre la evaluación externa o la evaluación conducida por la propia corporación docente.

Factores técnicos

Los estudios nacionales no reportan cuestionamientos importantes a la calidad de las pruebas, sino a la relación de sus características técnicas con un uso determinado. En este caso, se detectan dos controversias. En los países federales, existe una discusión acerca de en qué medida los contenidos que son evaluados en las pruebas representan una muestra confiable de las prescripciones curriculares subnacionales. La otra controversia detectada, en cambio, se refiere a la conveniencia, a las propiedades y a la pertinencia del uso de las evaluaciones censales y muestrales. Si bien existe una discusión técnica acerca de la validez de la información y de su uso según se la emplee para *high stakes* o para *low stakes*, la controversia también es parte de una discusión política acerca del modelo de implicancias de las evaluaciones. En este caso, el argumento sobre las limitaciones de las pruebas muestrales se extiende a las discusiones alrededor de las pruebas internacionales,

especialmente acerca del Programa para la Evaluación Internacional de Alumnos (PISA), pero también en relación con el uso de las pruebas nacionales. En algunos países incluso, los dispositivos asumen distinciones para diferentes usos y esta pareciera ser una tendencia que se extiende.

Los estudios destacan la asimetría en las capacidades técnicas de los estados subnacionales en los países federales y así emerge la importancia de fortalecer la capacidad técnica de los recursos humanos para gestionar las evaluaciones y sus usos.

El desfase en los ciclos de producción, difusión y uso de resultados también es señalado como una dificultad y constituye la razón que esgrimen los estados subnacionales al marcar limitaciones en el empleo durante el planeamiento de actividades con los establecimientos educativos.

Finalmente, en algunos países, se mencionan sospechas en determinados períodos respecto de filtraciones de pruebas que han minado la confiabilidad de los resultados.

Factores políticos

Los factores políticos que influyen en el uso –o no uso– de la información proporcionada por las evaluaciones tienen mayor o menor peso de acuerdo con la manera en que los países resuelven internamente las tensiones entre las funciones de la evaluación y según sus implicancias. Existen factores ideológicos que intervienen en las pujas entre actores, usualmente entre funcionarios y exfuncionarios gubernamentales cuando hay alternancia en el gobierno, entre gremios o actores que representan posiciones docentes –en particular ante usos con efectos regulatorios en la labor profesional–, entre funcionarios oficiales y también entre estos y las organizaciones de la sociedad civil, cuando tales organizaciones tienen posiciones tomadas en el debate sobre los modelos de evaluación y sus usos.

Asimismo, el espacio mediático también se convierte en caja de resonancia de los debates que, despojados de su complejidad técnica e impregnados a menudo de una lógica reduccionista, adquieren un peso difícil de eludir por los funcionarios con responsabilidades directas sobre los resultados de las políticas.

Los estudios permiten poner en relación los distintos tipos de factores y el modo como se configuran para favorecer o limitar los usos de los resultados de las evaluaciones en los procesos de toma de decisiones políticas. Posicionamientos políticos e ideológicos, conflictos entre actores en torno al uso de la evaluación para la gobernabilidad de los sistemas y el impacto mediático son factores que están presentes en la totalidad de los países. En cambio, lo que varía es la manera de procesarlos y lo que se postula en el presente informe es que el diseño institucional, la consistencia entre el instrumento de evaluación y el uso de sus resultados y la solvencia técnica de cada sistema permiten comprender la mayor o menor permeabilidad de los usos ante los factores políticos.

Factores comunicacionales

Los seis estudios en conjunto dan cuenta de los avances regionales en materia de apertura y diversificación de las políticas de difusión, si bien estos avances no son homogéneos y, en algunos países, se aprecian limitaciones en la continuidad o integralidad para la difusión de los datos.

El informe destaca la conveniencia de desarrollar políticas de difusión y comunicación continuas, previsibles, que periódicamente incorporen correcciones, actualizaciones y ampliaciones, como parte de una estrategia efectiva para consolidar la valoración de los resultados y construir su legitimidad. También se destaca la importancia de que la información sea concebida como un corpus integral compuesto por informes nacionales, subnacionales y sus distintos niveles

de agregación, presencia de microdatos en las plataformas interactivas con resultados de las pruebas y de cuestionarios complementarios y explicaciones metodológicas e informes de procesamientos especiales, todo lo cual contribuye a expandir la cultura del uso de la información.

Sin duda, el conjunto de los estudios pone de manifiesto la importancia de fortalecer la capacidad de comunicación y de procesamiento de los resultados en los estados subnacionales y en las unidades de gestión local en los países centralizados.

Por otra parte, entre los factores comunicacionales se señala especialmente en los informes la problemática de los medios. Por un lado, se constata que los medios, junto con las organizaciones de incidencia, constituyen un factor que contribuye a generar un debate público informado con base en los resultados de las evaluaciones y que este debate incide en el planeamiento de políticas, lo que ha sido también recogido en el inventario de los factores políticos. Al mismo tiempo, el modo reduccionista de tratamiento en algunos medios suma un argumento en favor del control y retrae la difusión de resultados por parte de los gobiernos, lo que también incide sobre el debate público. Este círculo vicioso pareciera poder moderarse con previsibilidad, continuidad en la difusión de resultados y con presencia plural de actores en el debate público.

En cuanto al uso de los resultados en el nivel de escuela, quizá no debiera concebirse únicamente como una problemática comunicacional, sino como parte de una política de formación de usuarios de los datos aportados por las evaluaciones.

Finalmente, el informe señala que es apropiado distinguir dos niveles de análisis al momento de incluir a los medios como un factor comunicacional en el uso de los resultados. Por un lado, el de observar una política de comunicación que incluya a los medios como usuario de la información y, al mismo

tiempo, como formador de opinión pública y actor presente en el debate. Por otro lado, se debe tomar en cuenta que diseñar una estrategia de comunicación en el marco de un programa de política determinado requiere tener una política de uso y no solo una estrategia de comunicación, ya que se trata de dos cosas diferentes.

El informe sugiere, en síntesis, la conveniencia de pensar en la comunicación como un proceso de construcción de usuarios en lugar de como un medio para alcanzar audiencias. También, se deberá comprender el papel de los medios como actores y no reducir las necesidades de interpretación y uso de la información de los actores a un problema de comunicación. En definitiva, se trata de situar la estrategia de comunicación en una política de uso de la información que vaya más allá de una estrategia de difusión de resultados.

Conclusiones

Los resultados de las evaluaciones contribuyen a generar diagnósticos que describen el comportamiento del sistema educativo, en función de niveles de logros académicos por sectores de población, identifican brechas en el cumplimiento del derecho a la educación o fortalecen hipótesis de intervención sobre unos y otros problemas. También constituyen la base para el establecimiento de metas de mediano plazo, aunque son menos frecuentes los reportes sobre el monitoreo de tales metas.

Los usos destinados a la coordinación entre actores nacionales y subnacionales han sido reportados con distintos grados de énfasis para el caso de los países federales. A la vez, la coordinación pareciera ser una condición cuya presencia o ausencia explica el uso o no uso de los resultados de las evaluaciones para la articulación de política y, al mismo tiempo, la fluidez de esa coordinación sería también explicada por el papel que la evidencia –de la cual los datos de la evaluación son una parte– cumple en el dispositivo de coordinación.

Se reportan usos más extendidos, aunque fragmentados en el empleo de los resultados de las evaluaciones a gran escala como insumos para el diseño y para la implementación de programas de mejora de logros sobre dominios específicos de conocimiento, siendo Lenguaje (según diferentes denominaciones en cada país) o Matemática los más comunes.

Con menor frecuencia se observa el empleo de los datos en programas de formación docente continua para enfocar la agenda de actualización profesional o para diseñar convocatorias a universidades o centros de formación que prestan servicios a la profesionalización docente. También se verifica su uso en la implementación del programa de formación, cuanto pueden constituirse en insumos para la preparación de recursos pedagógicos, como libros u otras herramientas destinados a la enseñanza.

También se identifican programas focalizados que seleccionan grupos de establecimientos a partir de indicadores de logro por debajo de un nivel aceptable, combinados con otros indicadores que también identifican condiciones desfavorables de otras variables de la escolarización. En estos casos, los datos proporcionados por las evaluaciones se integran a un conjunto de evidencias e informaciones más o menos estructuradas sobre problemas que ya se encuentran en agenda pedagógica, refuerzan diagnósticos previos y suman argumentos a decisiones ya tomadas o en proceso de construcción.

Por otra parte, el estudio no ha registrado entre los países que ha abarcado la práctica sistemática ni extendida de realizar evaluaciones de impacto, si bien se registra el uso de resultados de evaluaciones a gran escala para evaluar un programa de Bachillerato Internacional y la extensión de la jornada escolar en educación básica en dos países del estudio. Es probable que la baja frecuencia de uso para el monitoreo y la evaluación de programas se explique tanto por las características que asumen los datos de las evaluaciones –oportunidad, pertinencia, validez, cobertura– como por el hecho de que las

evaluaciones de impacto no forman parte de las prácticas generalizadas de análisis de políticas en los espacios estatales de producción de información.

En relación con los usos curriculares, la diferencia está marcada por la existencia o no de una política nacional de estándares, entendiendo por tal a aquella que establece parámetros de logro y su revisión periódica de acuerdo con los resultados obtenidos por las evaluaciones periódicas. En efecto, el empleo de resultados de evaluación y currículo depende del diseño de cada política curricular. Otros usos, como la revisión parcial de áreas específicas del currículo basadas en pruebas internacionales como el Programa para la Evaluación Internacional de Alumnos (PISA) o el Estudio Internacional de Civismo y Ciudadanía (CIVIC), no se apoyan estrictamente en resultados, sino en los modelos de organización de competencias que presentan las matrices de las pruebas.

Asimismo, se han registrado usos vinculados en forma más o menos directa con modelos de consecuencias fuertes, como el de incentivos salariales a docentes, brindar recursos a establecimientos condicionados a resultados y la publicación de *rankings*. Estas aplicaciones tienen una característica particular en relación con la formulación de políticas y es que se conforman como resultados que están al servicio de una política ya formulada, y que no necesitan de una mediación o de la construcción de alternativas para su implementación.

Recomendaciones

A continuación, se lista un conjunto de posibles acciones acerca del uso de resultados de las diferentes pruebas de evaluación de los aprendizajes.

- Anclar los usos de los resultados en acuerdos sobre las funciones de los sistemas de evaluación.
- Proyectar de manera progresiva los usos de las evaluaciones en el mediano plazo.

- Mantener la continuidad de las mediciones.
- Alinear los usos con las características técnicas de las pruebas.
- Definir políticas de uso de los resultados.
- Distinguir entre audiencias y usuarios.
- Fortalecer las capacidades técnicas de los estados sub-nacionales y de los usuarios locales.

1. Presentación del estudio

América Latina tiene una tradición de evaluación de aprendizajes de casi tres décadas, donde coexisten varios dispositivos¹ de evaluación ya consolidados junto con otros que están aún en proceso de formación. En el transcurso de estas tres décadas, hubo en la región varios replanteos acerca de las estrategias utilizadas para llevar a cabo las evaluaciones, motivados en parte por los avances técnicos que se han ido produciendo, pero también por las reacciones y los debates que se han ido generando entre diferentes actores de la sociedad. Actualmente, existe en la región un grupo de países que ha prescindido de las evaluaciones a gran escala. La discusión sobre las ventajas y desventajas de este tipo de instrumentos abarca desde temas ideológicos hasta prácticos, y cada vez emergen más voces que piden una revisión crítica de la teoría de cambio que vincula evaluación con mejora educativa.

A partir de la iniciativa del estudio *Use of Learning Assessment Data in the Planning Cycle* realizado por la sede del IIEP UNESCO, que pone el foco en los factores político-económicos que influyen en el uso de los resultados de las evaluaciones de gran escala en la región subsahariana de África, se planteó la conveniencia de realizar un estudio semejante en países de Latinoamérica con el doble propósito de generar un conocimiento contextualizado para los países de la región y para establecer comparaciones e intercambios de experiencias entre ambas regiones.

América Latina transita el desafío de lograr que los estudiantes mejoren sus niveles de aprendizajes. Según datos publicados por el Laboratorio Latinoamericano de Calidad

1 Se emplea la expresión “dispositivos” para enunciar a una variedad de modalidades institucionalizadas de evaluación.

Educativa (LLECE), la garantía del derecho a una educación de calidad para todos y todas aún no ha sido alcanzada en América Latina, y existen importantes desafíos en la adquisición de aprendizajes básicos, lo que afecta especialmente a los estudiantes más desfavorecidos (OREALC/UNESCO, 2015). Esta información es recogida por los dispositivos de evaluación, y es a través del análisis y el uso de esta información que se obtiene conocimiento susceptible de ser integrado y contextualizado considerando la trayectoria de las políticas educativas implementadas (cambios normativos, curriculares, ampliación de la cobertura, incremento de la inversión, etc.) y la identificación de factores que influyen en el aprendizaje (factores asociados). Quienes toman decisiones en los sistemas educativos pueden encontrar en el uso de esta información aportes interesantes para enriquecer el conocimiento e identificar áreas claves de intervención, basándose en evidencia empírica.

Si bien existen en la región sistemas de evaluación consolidados, maduros y con una sólida tradición en la producción de conocimiento, crece la discusión en torno a la subutilización que se hace de la información que estos dispositivos producen para fortalecer el vínculo entre el planeamiento de políticas y la mejora de una educación de calidad para todos los niños, niñas, adolescentes y jóvenes.

Desde este marco, cobra sentido elaborar un estudio que aborde los aspectos menos explorados que influyen en el proceso del uso de resultados de las evaluaciones de aprendizaje, enfocando no solamente los aspectos técnicos e institucionales, sino también las dinámicas y los elementos de política presentes en los distintos niveles de operación del sistema educativo. Tener mayor claridad sobre estos aspectos ampliará la conceptualización del problema y permitirá diseñar intervenciones con un aprovechamiento adecuado de la información que generan las evaluaciones de gran escala, a fin de alcanzar un mayor impacto en la mejora de la educación.

Objetivos del proyecto

El propósito del estudio es generar conocimiento específico y contextualizado para orientar el uso de los resultados de la evaluación de aprendizajes de gran escala de manera constructiva en el planeamiento estratégico y en la gestión de los sistemas educativos en distintos países de América Latina.

Específicamente, el estudio buscó indagar sobre la relación entre los usos de los datos de los operativos de evaluación de aprendizaje previstos en los marcos legales y en el diseño de los dispositivos, junto con su justificación y el uso efectivo que se hace de esos datos. La meta es generar conocimiento que permita un mayor aprovechamiento del potencial de estos instrumentos en el diseño de las políticas educativas de la región.

Para el logro de este objetivo general, se proponen los siguientes objetivos específicos.

1. Indagar sobre el uso de los resultados de los operativos de evaluación de aprendizajes comprometido por los organismos de planificación y gestión de la política educativa de cada país.
2. Identificar el uso efectivo que estos organismos hacen de los resultados de las evaluaciones en distintas fases del planeamiento de las políticas educativas.
3. Identificar qué factores influyen en los usos efectivos de los resultados de la evaluación para el planeamiento de políticas.

Para cumplir con estos objetivos, la Oficina para América Latina del IIEP UNESCO diseñó un estudio de seis casos nacionales y elaboró un informe comparado, que es exactamente el que se desarrolla en este documento.

2. Enfoque teórico y metodológico

A continuación, se presentan los principales ejes teóricos y metodológicos del estudio. En primer lugar, se sitúan cinco tópicos que forman parte del marco de trabajo y luego se describen los criterios de selección de la información primaria.

Las relaciones entre políticas públicas y planificación en América Latina, algunas reflexiones para determinar su encuadre

Para indagar en el uso de los resultados de la evaluación en las políticas educativas, es necesario plantear algunos interrogantes acerca del papel que cumple la planificación en la política educativa regional. Las reflexiones que se presentan a continuación procuran plantear cinco tópicos, con el fin de situar conceptualmente este interrogante.

Como reconocen los desarrollos del análisis de políticas públicas (Tamayo Sáez, 1997), pareciera necesario asumir un modelo de “racionalidad limitada” o “incremental” para la comprensión de los procesos de formulación de políticas. Desde esta perspectiva, se trata de considerar que la política educativa busca incidir en procesos sociales complejos en los que participan distintos actores, con capacidad de incidencia en los campos que se intentan regular u orientar. Asimismo, la naturaleza de las decisiones de gobierno implica que nunca es posible “comenzar de cero”, ya que siempre existen decisiones y compromisos previos, precedentes, cuyo desconocimiento, ruptura o reformulación, genera costos que los gobiernos se ocupan de medir a la hora de tomar decisiones. Estos costos exhiben valores diferenciales, sobre todo si el análisis se enfoca justo en el momento del recambio de gestiones de distinta orientación política, o si se refiere a países federales en los que el objeto de incidencia de la planificación

es la política educativa subnacional, es decir, bajo la gestión de identidades políticas diversas.

En segundo lugar y para la perspectiva regional, es importante tomar en cuenta algunos aspectos de la relación entre el Estado y los actores educativos. Durante la segunda mitad del siglo XX, el Estado fue, en América Latina, el artífice de la construcción de un espacio de cooperación dentro del cual tuvieron un lugar predominante los gremios, en tanto actores políticos y, en el sector público, particularmente los gremios educativos. Más adelante, los procesos de globalización y las reformas del Estado de finales del siglo XX reformulan parcialmente este consenso que había sido común a varios países de la región (Tiramonti, 2001). No obstante, en varios de estos, como Argentina, México o Uruguay, los docentes organizados todavía continúan manteniendo varios resortes de poder institucional dentro del Estado, un fuerte predicamento dentro del profesorado y una alta capacidad de movilización social, lo que los convierte en un actor importante en las coaliciones gobernantes o en las opositoras. Desde esta perspectiva, la voz gremial pareciera ser central en la construcción del sentido y la legitimación de la evaluación de resultados y, por lo tanto, de la legitimación y el sentido de su utilización.

Al mismo tiempo, la creciente participación de organizaciones de la sociedad civil provenientes del empresariado o de movimientos sociales en el debate educativo, así como la progresiva especialización de sus funciones con propósitos de incidencia, incorporan a la escena pública nuevos actores que hacen oír su voz en torno a la agenda educativa. A estas organizaciones deben sumarse las universidades, los organismos de investigación y los organismos internacionales que suman sus intervenciones públicas al debate.

El enfoque adoptado por este estudio asume que los dispositivos de evaluación integran el conjunto de herramientas de los organismos estatales para el ejercicio de las funciones de gobernanza de los sistemas educativos. En este sentido, también

se enfocará a los dispositivos de evaluación como un componente de la regulación del profesorado, que se encuentra en tensión por el control de la autonomía profesional y corporativa. Una corriente de la literatura que estudia los sistemas de evaluación en algunos países europeos y latinoamericanos analiza específicamente a la evaluación como parte de los dispositivos de regulación intraestatal que estarían desplazando sus modos de gobernanza desde los modelos burocrático-profesionales hacia modelos posburocráticos de gestión (Barroso y Carvalho, 2011; INEEd, s/f; Venegas, Dutercq y Maroy, 2017). Por este motivo, es posible considerar que los usos de la evaluación están influidos por las tensiones que despierta el proceso de modernización de la gobernanza del sector educativo en la región.

Otra dimensión de análisis que el estudio recorta se refiere al papel del diseño institucional de los dispositivos de evaluación y a su función en la definición de los usos previstos y efectivos de la evaluación. Regulados por ordenamientos legales que establecen compromisos entre los actores institucionales y frente a la sociedad, las orientaciones normativas serán objeto de una mirada que se encuentra en relación con los modos de institucionalización, entendidos estos como la manera en que distribuyen las responsabilidades entre las estructuras institucionales productoras y usuarias de información, junto con los márgenes de acción que otorgan a los actores frente a las prescripciones normativas.

El uso de la información proporcionada por las evaluaciones para abastecer con fundamento la toma de decisiones políticas, puede ser encuadrado en la problemática más abarcativa de la relación entre el conocimiento y las políticas. Desde esta perspectiva, la relación entre evidencia y decisión ha sido y es abordada desde diferentes enfoques. En nuestra región, viene siendo crecientemente caracterizada bajo el enfoque de “Políticas públicas basadas en la evidencia”. Esta noción, comprendida como el empleo de procedimientos sistemáticos para la construcción de conocimiento útil para la toma de decisiones (Bracho, 2010, entre otros), representa

principalmente una perspectiva normativa acerca de cómo debería ser la relación entre decisión y política, y tiene el propósito de fortalecer los procesos políticos mediante la elección de decisiones informadas y transparentes.

La tradición de los estudios del Policy Analysis ha enfocado históricamente esta relación desde una perspectiva que asume la complejidad de los procesos decisorios, caracterizándolos como parte de procesos de argumentación que construyen factibilidad ante la dificultad de alcanzar soluciones óptimas, dados los múltiples condicionamientos a los que están sometidas las decisiones políticas en los contextos democráticos (Majone, 1997). Este enfoque, asumido en el presente estudio, amplía los márgenes para el estudio del uso de la evidencia dado que está despojado de perspectiva normativa, si bien se encuentra anclado en una ética de la deliberación y del debate público propia de las sociedades democráticas (Lindblom, 1994).

Asimismo, Carol Weiss ha estudiado la relación entre conocimiento –bajo la forma de investigación y también evaluación– y política, y ha formulado tipologías que describen las múltiples caras que asume la complejidad entre evidencia y decisión política. Sus tipologías describen usos legitimadores, usos que amplían el conocimiento de los problemas, que contribuyen al debate público y a brindar usos instrumentales, estos últimos viables bajo condiciones de baja controversia, la sencillez de su aplicación y la estabilidad de sus entornos, condiciones que no siempre se cumplen en los escenarios políticos (Weiss, 1979 y 1998).

Los tópicos mencionados trazan las coordenadas principales del marco interpretativo general del estudio que se presenta. En adelante, se abordarán los procesos de decisión y desarrollo de las políticas bajo un modelo de análisis que asume la existencia de tensiones entre actores y contextos que limitan el alcance de las más óptimas decisiones. El marco interpretativo, por otra parte, recorta dentro de este enfoque

las relaciones entre los actores educativos institucionales y el Estado, al considerar a la evaluación –en sus múltiples formas– como parte de los dispositivos de regulación estatal que se incorporan a los debates públicos y como parte de los condicionantes que los decisores evalúan a la hora de tomar rumbos y definiciones. Asimismo, se asume que el diseño institucional de las evaluaciones enmarca la acción de los actores circunscribiendo el campo de tensiones mediante la prescripción u orientación de las conductas, mediante el establecimiento de mecanismos de coordinación y el trazado de márgenes de acción para el uso de la información en el campo de las decisiones. Finalmente, la problemática del uso de los resultados de las evaluaciones se ubica dentro de una temática mayor que es la relación entre el conocimiento y la decisión política.

Estrategia metodológica

El enfoque del estudio es de corte cualitativo y orientado a la práctica, para comprender los procesos que subyacen al uso observado de los resultados de las evaluaciones en distintos niveles del sistema educativo. Con este fin, también se considerará la comprensión, el significado y la experiencia de los distintos actores frente a las evaluaciones de gran escala. La metodología propuesta es el estudio de casos colectivo (Marradi, *et al.*, 2007), con énfasis sobre el componente comparativo intrarregional del estudio, pero sin descuidar los procesos involucrados en las trayectorias nacionales.

Efectivamente, se realizó un estudio en seis países de la región que cuentan con sistemas de evaluación consolidados e institucionalizados. Se trata de Argentina, Brasil, Chile, Ecuador, México y Uruguay. Estos seis países proporcionan una variedad de características que resultan apropiadas para la comparación. Tres países son federales, Argentina, Brasil y México. En tanto, los seis exhiben variadas trayectorias de institucionalización de sus dispositivos de evaluación y han definido diferentes modelos evaluativos desde el punto de

vista de las consecuencias que originan los resultados de las evaluaciones estandarizadas. Finalmente, también permiten acceder a perspectivas diversas respecto del uso que podría darse a las evaluaciones internacionales.

Para el logro del primer objetivo específico, se realizó un análisis de los compromisos asumidos acerca del uso de la información generada por los operativos de evaluación de aprendizajes, a través de una revisión de los marcos normativos vigentes, de los documentos oficiales de creación y difusión de los organismos y de los operativos de evaluación y los documentos de planeamiento de la política educativa, entre otros.

Para el segundo objetivo específico, se realizaron entrevistas semiestructuradas a actores clave del diseño e implementación de las políticas educativas: ministros y viceministros de educación, otras autoridades de primer nivel de las áreas de planeamiento, directivos de las instancias evaluadoras, equipos técnicos de gobierno, etc. Se entrevistó también a representantes de organizaciones de la sociedad civil, que son usuarias de la información proporcionada por las evaluaciones².

El tercer objetivo específico requirió un análisis de la información relevada en las dos etapas anteriores, tomando en consideración la apreciación específica que, sobre este tema, manifestaron los entrevistados, así como la opinión de los representantes de la sociedad civil, expertos, académicos y analistas, entre otros protagonistas consultados. Para este último fin, se realizó un conjunto de entrevistas que invitaron a desarrollar posibles hipótesis sobre las características que, en cada país, adquiere el uso de la información producida durante los operativos de evaluación de los aprendizajes en el planeamiento de la educación.

² Para obtener información más detallada acerca de los perfiles entrevistados, puede consultarse tanto el Anexo I de este documento como los Informes Nacionales.

El estudio abarca evaluaciones nacionales y evaluaciones internacionales. En el caso de las nacionales, se hace foco en aquellas que se aplican a la educación básica (CINE 1, 2 y 3). En cuanto a las internacionales, se indagará en todos los países la implementación de las pruebas PISA y las que aplica el Laboratorio Latinoamericano de Evaluación de la Calidad Educativa (LLECE). En los casos de Chile y México, que además de las mencionadas aplican también otras evaluaciones internacionales, se incluyeron aquellas que surgieron como más relevantes en los testimonios de los entrevistados.

El estudio cubre distintos periodos, en cada uno de los seis países, aunque todos pueden incluirse en la denominada “tercera oleada”, según la periodización propuesta por Cecilia Galas para trazar la trayectoria de los dispositivos de evaluación en la región (Galas *et al.*, 2020)³. No obstante, las fechas deben ser tomadas solo como una referencia, dado que los informes se remontan a años previos en aquellos casos en los que es relevante hacerlo o cuando los entrevistados se refieren a operativos de evaluación anteriores, para la reconstrucción del uso en ciertas políticas o para la descripción de antecedentes.

Acerca de este informe

A continuación, se presenta el informe comparado a partir de la información recabada y luego procesada en cada uno de los casos nacionales abarcados por el estudio. El propósito es desarrollar una reflexión general e identificar y proponer

³ De acuerdo con la autora, en esta oleada se incorporan dominios no cognitivos en las evaluaciones; se desarrollan índices complejos en los que el desempeño académico se integra como un componente más junto con indicadores de trayectoria; se extiende el uso de los cuadernillos complementarios incorporando información proveniente de docentes y familias; se avanza en la segmentación de las audiencias para comunicar resultados; se verifican variaciones en los formatos de los reportes; se asiste al desarrollo de herramientas para la visualización y la obtención de procesamientos simples con formatos apropiados para el uso de los centros educativos, docentes, padres y medios de comunicación; se registra la incorporación de sistemas de consulta y presentaciones interactivos, y se avanza en la generación de modalidades de evaluación para usos pedagógicos al incorporar pruebas que permiten que los docentes registren los avances de sus estudiantes durante el año (Galas *et al.*, 2020: pp. 18-19).

interpretaciones sobre las racionalidades que pueden advertirse detrás de los usos de las evaluaciones estandarizadas para las políticas educativas. Dado que se trata de un texto con propósitos interpretativos, la presentación de la información se limitará a la necesaria para el desarrollo de la argumentación que orienta su narrativa. Por tanto, el lector puede remitirse a los Informes Nacionales cada vez que desee profundizar la información específica de cada país.

3. Trayectorias comparadas de los dispositivos de evaluación de aprendizajes a gran escala abarcadas por el estudio. Un marco de interpretación

En este capítulo se despliega un marco de interpretación para encuadrar el análisis del uso de la información de las evaluaciones estandarizadas. El marco se organiza en torno a tres ideas fuerza. Por un lado, se considera que existen dos funciones generales atribuibles a los dispositivos de evaluación y que, más allá de los énfasis con que los países adopten una u otra función, ambas se encuentran en tensión. En segundo lugar, se plantea la idea de que las tensiones mencionadas y su reflejo en el diseño institucional moldean los usos de la evaluación para el diseño de políticas. Por último, se señala que pueden advertirse antecedentes acerca de estas tensiones, en el origen de los dispositivos de evaluación.

En la siguiente sección, se presentan las tensiones descritas y se despliega el resto de los componentes del marco interpretativo mediante la reconstrucción de las trayectorias de los dispositivos de evaluación. Luego, en una segunda sección, se van introduciendo los argumentos a medida que se describen las trayectorias de los dispositivos específicamente para el caso de cada uno de los países.

Evaluaciones estandarizadas: dos funciones en tensión

Los sistemas de evaluación suelen describirse sobre la base de dos funciones amplias de acuerdo con las implicancias que tienen sobre los actores. La literatura anglosajona habla de *high stakes assessment* y de *low stakes assessment* para

categorizar evaluaciones con altas o bajas consecuencias sobre los actores, sean estos estudiantes, docentes, instituciones, autoridades educativas o programas (Ravela, 2006). Aunque esta clasificación simplifica la complejidad de las funciones de las evaluaciones estandarizadas, es útil como un modelo de interpretación que establece un marco amplio para la comprensión de los usos de las evaluaciones en las políticas educativas. Ravela (2006) asocia los usos con implicancias altas a lo que comúnmente se denomina evaluaciones sumativas, en tanto asocia los usos de implicancias bajas a lo que se denomina usos diagnósticos, formativos o para la mejora.

Sin embargo, más que un modelo de clasificación binario, el esquema de implicancias altas o bajas permite leer en cada caso la coexistencia, el predominio o las tensiones entre ciertas funciones polares de la evaluación. Algunas funciones tienen consecuencias para los individuos, pero no necesariamente caracterizan al sistema, como es el uso de algunas pruebas para la certificación de estudios que están orientadas a la evaluación de las personas y no a la de los sistemas. La función de ordenamiento según resultados puede tener implicancias directas sobre los individuos cuando el orden de mérito implica el acceso a algún tipo de beneficio o el ingreso en un nivel educativo superior, en cambio, puede tener consecuencias indirectas cuando el orden de establecimientos es una forma de entregar información sobre el desempeño de las instituciones educativas al mercado. El empleo de incentivos puede tener implicancias directas sobre el salario de los docentes o sobre el acceso de las escuelas a determinados beneficios. La rendición de cuentas puede tener implicancias confrontativas cuando su uso es predominantemente el de “echar culpas” sobre actores mediante la responsabilización unilateral o, en cambio, usos cooperativos cuando su empleo se orienta específicamente a identificar y monitorear las responsabilidades compartidas de los distintos agentes, incluyendo a los agentes estatales, de la sociedad civil y a los movimientos sociales, si están al servicio de espacios deliberativos (UNESCO, 2017).

La utilidad de la versión binaria de esta clasificación, a pesar de ser simplificadora, es que permite visualizar el lugar que tiene el sistema y el proceso de decisiones en el uso de los resultados. En las evaluaciones de altas implicancias, el resultado está vinculado a una decisión de manera más o menos directa o automática y, en algunos casos, podría afirmarse que preexiste al resultado. Por ejemplo, los sistemas que prevén sanciones a establecimientos que no cumplen con determinados estándares, la asignación de incentivos salariales a los profesores de acuerdo con el rendimiento académico de los estudiantes, la distribución de recursos asociados al cumplimiento de determinados resultados, la certificación de un estudiante en función del puntaje obtenido en una prueba de finalización del nivel o el ordenamiento de personas o de instituciones para la selección. En otros casos, como el de la publicación de *rankings*, tienen efectos de sanción indirectos, a través de mecanismos de mercado –esto ocurre cuando existen posibilidades de optar por el servicio–, o a través de efectos de legitimación. En todos estos casos, existe un criterio que “dispara” una decisión o un procedimiento.

En los usos diagnósticos o formativos, la decisión tiene una relación mediada con el resultado por algún tipo de proceso, es decir que no es inmediata. El diagnóstico ilumina un camino, orienta una decisión, permite seleccionar una opción entre varias, pero el siguiente paso no se deriva de manera “automática”; no es común que una decisión política se “dispare” de un diagnóstico, sino que, como se sabe –y el enfoque de planeamiento de situaciones lo ha puesto de manifiesto– es parte de un proceso de diseño, de deliberación y de construcción para dar viabilidad a la decisión. Como se ha manifestado desde el enfoque conceptual de este estudio, el uso de la información para la construcción de viabilidad (Majone, 1997) ha sido registrado desde hace décadas y hoy bien puede identificarse como uno de los usos de la evaluación formativa. Los usos formativos o para la mejora requieren, asimismo, de la construcción de soluciones o alternativas que tampoco están contenidas en el resultado, sino que en muchos casos se exige

un análisis complejo de la información. Una política de mejora no se deriva de manera directa de una serie de resultados de evaluación: aunque estos ayuden a precisar el problema, la solución deriva de una construcción, tanto política como técnica, y requiere de procesos de deliberación.

Desde la perspectiva asumida por este estudio, las funciones de altas implicancias y de bajas implicancias pueden verse como sendas modalidades de regulación con que los países resuelven la gobernabilidad de sus sistemas educativos (Venegas, Dutercq y Maroy, 2017; Barroso y Carvalho, 2011). Como se ha señalado en el enfoque teórico, las pruebas de evaluación participan del gobierno de los sistemas educativos en el sentido de constituirse como un insumo para que las autoridades y los actores orienten su accionar.

De este modo, es posible afirmar que, más allá de paradigmas o funciones predominantes en cada uno de los países, ambas funciones pueden ser visualizadas en los distintos casos, ya sea a través de los usos previstos o efectivos de los dispositivos como de las posiciones de los actores al valorar los usos de la información.

Las funciones de implicancias altas o de bajas implicancias expresan preferencias de los actores con capacidad de incidencia sobre los sistemas educativos y sobre los dispositivos de evaluación, en una puja por la gobernabilidad de la educación. Las funciones que se ejercen en el marco de ambos modelos algunas veces se solapan o entran en tensión, pero, al mismo tiempo, constituyen la manifestación de visiones y perspectivas diferentes de los actores sobre la evaluación. La coexistencia o la tensión modelan los usos generales de la información y, junto con otros factores que se describirán en este estudio, explican los usos específicos orientados hacia la toma de decisiones de política.

A lo largo del informe, se hará referencia a la capacidad de los diseños institucionales de algunos países para absorber

tensiones. Esto significa que algunos diseños institucionales exhiben cierta capacidad de incorporar de manera progresiva –y en el marco de debates muchas veces ásperos– funciones que logran moderar la definición dominante.

Una interpretación de la trayectoria de los dispositivos de evaluación en los seis países estudiados

En el origen de los dispositivos de evaluación hay reformas al diseño institucional estatal. En cuatro de los países bajo estudio, las reformas constitucionales son parte del proceso de rediseño del Estado (Brasil, Chile, Ecuador y México) y en todos se asiste a la renovación de los sistemas legales en el sector educativo. En el caso de Argentina, Brasil y México, la introducción o el refuerzo de las evaluaciones estandarizadas a gran escala están asociados con los procesos de descentralización de los sistemas educativos, como ya ha señalado la literatura y como también se recoge en este estudio. En los países con sistemas centralizados, también hay reformas educativas de envergadura. En Chile, por ejemplo, la evaluación se introduce en el marco del rediseño de su sistema educativo y en particular del mecanismo de financiamiento de la educación. En Ecuador, se afianza en el marco de una reforma integral y bajo el establecimiento de un plan estratégico a 10 años (2006-2015), que se propone incrementar la cobertura del sistema, establece el incremento del presupuesto en educación, y se propone impulsar el monitoreo y la mejora de la calidad y la equidad con el uso de la evaluación como herramienta. Uruguay, por otra parte, presenta una trayectoria condicionada por una definición muy fuerte en favor de la evaluación formativa, en cuyo desarrollo y aplicación tienen un papel central los estamentos docentes con representación en el gobierno de la educación.

Tener presente este origen es importante, porque la función o las funciones asignadas a la evaluación estandarizada en los momentos de constitución de los dispositivos de evaluación, así como la trayectoria de estabilidad y consolidación, de revisión y de resolución de tensiones internas entre las distintas

funciones y actores que las apoyan, constituyen ejes para la comprensión de los usos que luego se da a los resultados para la toma de decisiones en política educativa. En este sentido, las trayectorias de los seis países en materia de evaluación no son lineales, es decir que la adopción de tendencias globales no impacta de la misma manera en los sistemas educativos y las tensiones, marchas y contramarchas que algunos países exhiben son indicadores de las maneras de apropiación de los flujos globales de política. Es importante, por lo tanto, remarcar el contexto político de surgimiento de las evaluaciones a gran escala y el peso que adquieren los debates académicos y públicos sobre las funciones regulatorias que estos dispositivos tendrían sobre los sistemas educativos en procesos de reformas, sobre la burocracia educativa y sobre el disciplinamiento del cuerpo docente (Barroso y Carvalho, 2011; Maroy, 2008).

A continuación, se toman para el análisis los tres países federales. Así se observa que en la trayectoria del sistema de evaluación de Brasil se destaca un proceso de institucionalización asentado en la interacción y el mutuo refuerzo de tres ejes. Por un lado, está el diseño institucional del federalismo educativo, anclado en una fuerte base de distribución de recursos de impuestos federales, estatales y locales. Este eje permite seguir la trayectoria de consolidación del Sistema de Evaluación de la Educación Básica (SAEB, por sus siglas en portugués), cuya función es producir parte de la información sobre la cual se organiza una fracción de la transferencia de recursos no obligatoria de la Unión a los entes federados y, a partir de 2023, sobre una fracción de los repases obligatorios. Dada esta característica, dicho sistema se constituye en un referente con el que se alinean las evaluaciones subnacionales y esto tendrá como consecuencia que las iniciativas estatales y municipales autónomas de evaluación existentes también se sintonicen con el SAEB.

El segundo eje es la adecuación de las funciones del INEP y su constitución como organismo evaluador (función que se agrega a las amplias y reconocidas funciones que ya tiene

como productor de información e investigaciones sobre educación). Finalmente, la construcción de un Índice de Educación Básica (IDEB, por sus siglas en portugués) en 2007 aceleró la conformación de sistemas de evaluación estaduales. La existencia de programas de transferencias de recursos de la federación hacia los estados y municipios asociados con el IDEB también serán parte de este proceso. Como podrá apreciarse a lo largo del estudio, el diseño federal explica gran parte del uso de la evaluación en los niveles decisorios federales y subnacionales.

En el caso de México, la evaluación estandarizada a gran escala también está asociada con la descentralización. Si bien el país ya tenía antecedentes tempranos (que se remontan a finales de la década de los sesenta y a la década de los ochenta), que están asociados a la regulación del flujo de matrículas para el ingreso a la universidad y al nivel secundario, es a partir de 1992, cuando se firma el Acuerdo Nacional para la Modernización de la Educación Básica (ANMEB) y, a partir de 1993, cuando se reforma el artículo 3° del Acuerdo Constitucional y se sanciona la Ley General de Educación. Estos marcos establecen la descentralización y la asignación de las funciones de la evaluación a la Secretaría de Educación Pública (SEP), instalan las evaluaciones y de este modo se inicia el proceso de fortalecimiento de las áreas estatales de evaluación.

En este país, la trayectoria de las evaluaciones puede ser seguida durante una primera etapa en la que se fortalecen las capacidades técnicas de los organismos de evaluación y se despliegan distintas iniciativas de evaluación a gran escala (EXALE, ENLACE, PLANEA), así como a través de las tensiones internas entre gobierno y sindicato, motivadas por el papel de estas iniciativas en la regulación de la carrera docente y por las tensiones intraburocráticas entre la SEP y el Instituto Nacional de Evaluación de la Educación (INEE), que fue creado en 2002, alcanzó su autonomía en 2013 y luego fue cancelado en 2018, cuando el gobierno decidió revisar radicalmente el sistema de evaluación.

México ilustra con bastante claridad la coexistencia y la tensión de las funciones de altas y bajas implicancias. Por un lado, empleó la evaluación de resultados como parte de un esquema organizador de la carrera magisterial y de distribución de incentivos hacia el cuerpo docente, lo que permite afirmar que el sistema desarrolló un componente del modelo de implicancias altas. Pero a la vez contó con tres modalidades de aplicación de la evaluación estandarizada PLANEA, algunas de las cuales estuvieron orientadas a la función formativa⁴. Este esquema de tres modalidades de aplicación de las evaluaciones suponía un balance de la función de evaluación externa propia, aunque no exclusiva, del modelo de implicancias altas y del modelo de evaluación formativa, ya que PLANEA ELCE y PLANEA Diagnóstica cumplían con esta función.

En Argentina, la evaluación muestra una trayectoria de consolidación en cuanto a su permanencia y complejización técnica, que se asocia con un federalismo de concertación consolidado mediante su Consejo Federal de Educación, aunque sin mecanismos que liguen distribución de recursos con resultados de evaluación, como ocurre en el caso de Brasil. Tal como se retomará más adelante, el Consejo Federal de Educación formula planes de evaluación, con objetivos amplios y sin implicancias altas. Los vaivenes de la trayectoria de las evaluaciones que registra este país parecieran estar vinculados con la discusión acerca de las funciones de la evaluación y con la noción de calidad de la educación. En Argentina, las controversias pasan por la construcción de un modelo de evaluación que asuma una noción amplia o una noción estrecha de la calidad de la educación, según los períodos, y pareciera que la no resolución de esta controversia posterga la definición de los usos de los resultados sobre la

4 PLANEA ELSEN (a cargo del INEE) evaluaba el desempeño del Sistema Educativo Nacional a partir de los logros en la educación obligatoria; PLANEA ELCE (a cargo de la SEP) tenía el propósito de retroalimentar a las comunidades escolares de primaria, secundaria y de educación media superior, y PLANEA Diagnóstica (aplicada y calificada por los docentes de aula) tenía el objeto de evaluar aprendizajes clave de Lenguaje y Comunicación y de Matemáticas a estudiantes del cuarto grado de primaria, al inicio de cada ciclo escolar.

macropolítica, lo que no impide, como podrá observarse, el despliegue de ciertos usos específicos en distintos niveles del sistema y en las provincias.

En cuanto a los países con gobiernos de educación centralizados, una mirada general a su trayectoria expone lo siguiente. Uruguay asume tempranamente los propósitos formativos como la definición central de su política de evaluación. Esta definición inicial por un uso de los resultados sin altas implicancias, así como el desarrollo de una estrategia incremental para potenciar el uso de las evaluaciones y de sus resultados por los docentes en el aula, constituyen los ejes que permiten caracterizar la trayectoria de su sistema de evaluación. Mantiene la confidencialidad de los resultados, excluye a los *rankings* como opción de uso de la información, ha desarrollado dispositivos para el uso formativo y ha incorporado Aristas en Clases y la Evaluación Adaptativa, para reforzar la función formativa. Esto implica que es posible afirmar que Uruguay resuelve las tensiones que se verifican durante la instalación de los dispositivos de evaluación en varios de los países estudiados por medio de la construcción de un amplio consenso alrededor de la función formativa para su uso por parte del cuerpo magisterial.

La definición por el uso formativo en el micronivel no excluye, no obstante, ejercer otras funciones de la evaluación, como lo son las de diagnóstico en el nivel macro. Al mismo tiempo, desde mediados de la década pasada, se vislumbra que a partir de la creación del autónomo Instituto Nacional de Evaluación Educativa (INEEd) –que ha desplegado sus propias evaluaciones–, se habilitó un debate creciente acerca de incorporar usos vinculados con el monitoreo externo y con la responsabilización por los resultados. En tanto, la tensión con la línea de fundación original pareciera apreciarse en la convivencia entre un sistema de evaluación bajo de responsabilidad, en el caso de la Administración Nacional de Educación Pública (ANEP), que es heredera del uso formativo, y el caso del INEEEd, que tensiona hacia funciones de evaluación,

monitoreo y responsabilización, según lo destacan los funcionarios que han sido entrevistados.

Chile, por otra parte, tiene una trayectoria comparable a Brasil en términos de fortalecimiento institucional, aunque presenta un motor diferente del de la coordinación federal, que ha contribuido a organizar su sistema. Este país ha desarrollado un diseño basado en consecuencias fuertes para regular un sistema educativo que funciona bajo el modelo de cuasimercado, que también asume la responsabilidad de apuntalar los establecimientos que no alcanzan los niveles mínimos de rendimiento. El diseño del Sistema de Aseguramiento de la Calidad (SAC) condiciona el margen de acción del nivel macropolítico y deja en manos de los niveles meso y micro la responsabilidad por las acciones para la mejora educativa. Asimismo, en comparación con Uruguay, Chile ha seguido la trayectoria inversa en cuanto al balance de sus funciones y, a partir de un diseño orientado a la responsabilización, ha ido ajustando su esquema sin perder su impronta original, para incorporar tempranamente la función de producir información dirigida al apoyo y la mejora de los establecimientos. Mientras que Uruguay en la etapa actual del desarrollo de sus evaluaciones está experimentando tensiones internas para revisar la hegemonía de la función formativa y algunos actores señalan la necesidad de incrementar las funciones orientadas al monitoreo y la responsabilización, Chile ha ido experimentando tensiones inversas debidas a los efectos que su sistema genera sobre las escuelas y en sus actores educativos, lo que ha movilizado a otros protagonistas de la sociedad civil organizada. Aún están por verse las consecuencias que tendrá sobre el sistema de evaluación de este país el proceso de reforma constitucional que se ha abierto al calor de las protestas sociales. De manera similar, en Brasil existe un creciente interrogante de parte de los funcionarios y de los usuarios de las organizaciones de la sociedad civil acerca del fortalecimiento de la función formativa, así como algunas iniciativas subnacionales que van en la misma dirección.

Entre los casos estudiados, Ecuador se constituye como el país de más reciente instalación de un sistema de evaluación a gran escala. Si bien posee antecedentes desde mediados de los años noventa, cuando realiza evaluaciones muestrales en el marco de un proyecto para sectores urbanos vulnerables con apoyo internacional, es a partir de 2009 cuando encara la instalación de un Sistema de Evaluación y Rendición de Cuentas (SER) orientado –como su nombre lo indica– a la medición de logros y a la responsabilización por los resultados. El Sistema, producto del mandato de una reforma constitucional realizada en 2008 y en el marco de un Plan Decenal, sancionado en 2006, y orientado principalmente a incrementar la cobertura y a aumentar el presupuesto educativo, también incluía la evaluación de la gestión institucional, el desempeño de los docentes, directivos y estudiantes, y la calidad del currículo.

En este país, la trayectoria de las evaluaciones, que verifica como en el resto de la región la incorporación progresiva del estudio de factores asociados y que incorpora en 2012 al Instituto Nacional de Evaluación (INEVAL) como organismo autónomo, pareciera exponer como rasgo predominante una tendencia diferente entre el tratamiento dado a su educación básica (CINE 1 y 2) y a la secundaria superior (CINE 3), representada por la aplicación de las pruebas Ser Estudiante y Ser Bachiller, respectivamente. Es posible observar que las primeras muestran una continuidad de propósitos si bien se registran discontinuidades en la frecuencia de publicación de los resultados, mientras que las segundas parecieran haber experimentado vaivenes en relación con su función, pues se originan con un doble propósito, ya que además de emplearse para informar sobre los niveles de logro alcanzados por la población escolar que asiste al bachillerato, se constituyen como un examen de certificación de finalización de estudios. Esta función de regulación del flujo de matrícula, se acentúa incluso en 2014 y en 2017, cuando los resultados pasan a formar parte de los requisitos de puntuación para el ingreso a la educación superior, primero mediante su aprobación y luego como un porcentaje del puntaje requerido para el

ingreso. Un poco más tarde, en 2020, Ecuador informa haber desdoblado estas funciones en dos evaluaciones, asignando al INEVAL la función de aplicar una evaluación como examen de grado y al organismo pertinente de la educación superior, la responsabilidad de una evaluación específica para determinar el ingreso a ese nivel por parte de los egresados del bachillerato.

La dinámica de la globalización de políticas en un campo especializado, intensivo en intercambios académicos e institucionales, sumamente internacionalizado –con la presencia de jugadores globales de alta *performance* como lo son los organismos internacionales– explica las tendencias comunes de la evaluación a nivel regional. Situar tensiones internas (algunas de las cuales estuvieron presentes en los momentos fundacionales) y rastrear su desarrollo a lo largo de las décadas permite explicar la manera en que los países asimilan las tendencias globales en función de sus propias dimensiones. Ciertamente, sería necesario un estudio bastante más específico para trazar en detalle cuál es el despliegue de estas tensiones, no obstante, el panorama general presentado en las páginas anteriores y los detalles que se leen en los seis estudios regionales proporcionan claves para sostener esta explicación como soporte interpretativo destinado al análisis de los usos de las evaluaciones.

En síntesis, cinco países han logrado estabilizar las tensiones entre las dos funciones polares de los modelos de evaluación: Brasil, Chile, Ecuador y Uruguay lo han alcanzado con firmes definiciones iniciales; en tanto Argentina, con algunos vaivenes institucionales e indefiniciones en los usos, muestra una estabilidad construida a costa de no haber zanjado aún una discusión histórica sobre la vinculación entre los resultados de la evaluación y las políticas acerca de la calidad educativa. Por otra parte, al momento de la realización del presente estudio, México se encontraba en un proceso de reformulación de sus evaluaciones y de su diseño institucional, como fruto de una eclosión de tensiones internas alrededor de las

funciones de la evaluación. En definitiva, la modalidad con que cada país ha alcanzado la estabilización pareciera tener consecuencias sobre el papel que cumplen las evaluaciones en sus respectivos sistemas educativos y, consecuentemente, en el uso que se hace de los resultados.

4. Usos previstos de las evaluaciones estandarizadas según los marcos normativos

Este capítulo busca identificar y caracterizar cuáles son los usos previstos de las evaluaciones para el planeamiento de las políticas educativas en los diferentes marcos normativos nacionales. Se vuelve evidente que los usos previstos para esta función específica no surgen de una lectura lineal de los enunciados normativos, dado que un abordaje de este tipo supone una incidencia directa de la letra de los marcos legales en la conducta de los actores. El enfoque que se asume, entonces, es el de una perspectiva holística que lee, además de los propósitos y usos enunciados, cuáles son los mandatos regulatorios del sistema de evaluación en su conjunto, bajo el supuesto de que los usos (o no usos) para el planeamiento de la política están enmarcados en las funciones dominantes del modelo de evaluación, moldeados por la dimensión estatutaria⁵ y también por la trama normativa que integran.

Funciones asignadas por los encuadres normativos

Para identificar las funciones asignadas a la evaluación, inicialmente se revisaron los enunciados de la normativa cuando se refieren a propósitos o a funciones. En este sentido, todos los países, aunque con distintos énfasis y especificaciones –tal como se verá– asignan las siguientes funciones que se derivan de enunciados genéricos:

- asegurar y mejorar la calidad de los aprendizajes;
- Contar con información para la toma de decisiones informadas (diagnósticos, monitoreo, evaluación de políticas);

⁵ Se refiere a la dimensión reglamentaria de una norma, es decir, aquella que establece los procedimientos específicos, las acciones y las conductas para los sujetos que mandata.

- verificar el grado de cumplimiento de las metas de aprendizaje;
- Coordinar entre actores subnacionales;
- proporcionar información pública para la responsabilización sobre los resultados; transparencia;
- certificar la finalización de estudios de un nivel educativo;
- promover una cultura de la evaluación (por ejemplo, propiciar competencias evaluativas en los establecimientos).

La amplitud de la enunciación se explica porque el discurso normativo posee funciones argumentativas además de las prescriptivas, sin embargo, para los objetivos de este estudio es necesario no sobreestimar las primeras e identificar las segundas. Por lo tanto, los enunciados sobre las funciones serán analizados en relación con el modelo de implicancias predominante y con el tipo de prescripción que establecen a los actores, para cumplir con tales funciones.

Encuadres normativos y usos previstos en los países del estudio

La Ley de Educación Nacional (LEN) de Argentina asigna al Ministerio de Educación Nacional la responsabilidad sobre la evaluación y enuncia sus características generales. Este marco legal es complementado por los Acuerdos Federales que indican que la Nación y las provincias deben crear espacios de concertación y de planificación, como la Red Federal de Mejora de los Aprendizajes, creada en 2016, “para planificar acciones de mejora a través del trabajo articulado y colaborativo entre los equipos nacionales y provinciales” (Resolución CFE 284/06 Anexo I), en tanto, a partir de 2019 se institucionaliza la Red Federal de Evaluación. En el nivel nacional, estos marcos se complementan con las misiones y funciones de la Secretaría de Evaluación e Información Educativa que tiene asignada la responsabilidad técnica de producir información.

La mejora de la calidad es un enunciado presente en todas las normativas y documentos de orientación del país. Su Ley

de Educación Nacional establece que es responsabilidad del Ministerio de Educación en acuerdo con el Consejo Federal de Educación implementar una política de evaluación para la mejora de la calidad. En el capítulo sobre trayectorias comparadas, el estudio sobre la Argentina identifica una tensión entre una noción ampliada y una noción restringida de calidad en la que se evidencia una preocupación por el cumplimiento de los derechos. Esto también se recoge en los marcos normativos. En la Resolución 549 de creación del Consejo Nacional de Calidad, se enuncia que el Consejo deberá asumir “... una visión amplia de la calidad de la educación... (e) ir más allá de los debates centrados únicamente en los aprendizajes medibles en operativos de evaluación y generar una perspectiva integral que redefina el concepto de calidad de la educación ...” (Resolución 549/20, Creación del Consejo Nacional de Calidad). Como puede leerse, pareciera ser que el Consejo Nacional de la Calidad tuviera asignada la tarea de proponer una solución a un debate a cuya resolución está sometida la función principal del dispositivo de evaluación⁶.

En cuanto a la realización de diagnósticos, monitoreo y evaluación de políticas, ya se ha mencionado que la LEN asigna al Ministerio de Educación la responsabilidad para el “desarrollo e implementación de una política de información y evaluación continua y periódica del sistema educativo para la toma de decisiones (...)”. También se establece que se asigna a la Secretaría de Evaluación e Información Educativa (SEIE) la responsabilidad de proveer insumos estratégicos para el diseño, el monitoreo y la evaluación de políticas educativas y, finalmente, se establece que el Consejo Federal debe contar con diagnósticos y monitoreo, para lo cual se constituye una Red integrada por las autoridades nacionales y provinciales, los equipos de supervisión/inspección y de gestión escolar y los docentes (Consejo Federal de Educación, Res. 284, Anexo I).

6 Aunque su creación estaba prevista desde 2006, recién fue creado en 2020.

La función de diagnosticar, monitorear y evaluar es asignada a un conjunto de agentes de distintos niveles nacionales, provinciales, de nivel meso y de nivel escolar. Se trata de una función extendida que, al menos en este modo de enunciación, supone un desempeño autónomo y exige un alto nivel de coordinación.

En cuanto a la función de coordinación entre estados subnacionales, Argentina, junto con Brasil exhiben enunciados específicos orientados a la coordinación federal, si bien lo hacen mediante modelos diferentes. El de Argentina, puede ser caracterizado como un modelo abierto o de consenso, dado que el marco normativo define un espacio de mediación, de concertación, de colaboración y planificación, bajo la figura espacios interjurisdiccionales, como lo es el de la Red de Mejora de los Aprendizajes. En esta línea, el marco normativo no establece pautas o criterios de coordinación, sino que las deja abiertas para que sean definidas en espacios de negociación, para lo cual las evaluaciones conforman un insumo.

Por otra parte, la normativa nacional sobre evaluación de la Argentina no establece entre sus funciones la verificación del cumplimiento de metas de aprendizaje, más allá de alguna mención general acerca del cumplimiento de los aprendizajes esperados y más allá de una iniciativa, que luego fue discontinuada, de promover una herramienta para la planificación (POA-I) en 2016, que procuraba asociar acciones de política con resultados de evaluaciones.

A diferencia de lo que ocurre en Brasil, Chile o Ecuador, las políticas curriculares de la Argentina no se orientan por la fijación y la evaluación de estándares, sino por marcos curriculares nacionales que se plasman en Acuerdos Federales y provinciales, que proponen logros que no alcanzan la fuerza regulatoria de los estándares o de la Base Nacional Común Curricular de Brasil. Aunque en 2018 se aprobaron los “Indicadores de Progresión de los Aprendizajes Prioritarios de Matemática” (CFE 342/18), lo que puede entenderse como un intento por especificar el marco curricular de esta asignatura, el estudio

nacional no arroja evidencia de que estos indicadores hayan logrado permear la política de evaluación, ni la política curricular.

Aunque no sea una función establecida por la normativa de nivel nacional, los dos estados subnacionales de la Argentina que se abordan en este estudio demuestran su preocupación por obtener algún tipo de retroalimentación, sino es del grado de cumplimiento, al menos sobre los aprendizajes prescriptos en sus marcos curriculares mediante una selección de contenidos. En el caso de la Ciudad Autónoma de Buenos Aires (CABA), sus evaluaciones de Finalización de Estudios Primarios (FEPBA) permiten obtener información sobre “algunos” de los logros de aprendizaje establecidos en el currículo de la CABA para el nivel primario. Por otra parte, las pruebas de Finalización de Estudios Secundarios (FESBA) han sido diseñadas para obtener información sobre “algunos” de los logros de aprendizaje establecidos en los contenidos curriculares para el nivel secundario y en las metas de aprendizaje (las comillas resaltan el textual de los testimonios relevados por el estudio nacional). Mientras que, en la provincia de Córdoba, se ha desarrollado un sistema de evaluación provincial para brindar anclaje y referencias a los diseños curriculares locales.

En cuanto a la responsabilización, la Argentina establece la obligación de informar al Poder Legislativo. De este modo, la LEN 26.206 de 2006 establece que: “Es obligación del poder ejecutivo elevar anualmente un informe al Congreso de la Nación dando cuenta de la información relevada (art. 99)”, si bien –como se observa– no se presenta una referencia específica a la evaluación sino a la información en general, en tanto el artículo 94 de la LEN, hace referencia a “una política de evaluación para (...) la transparencia y la participación social”. De hecho, un informe de gestión como obligación para la rendición de cuentas corresponde a una función general de la evaluación sin implicancias directas, sino solo con usos formativos. Esta afirmación, aunque sea obvia, encuadra una modalidad de uso de la información para la toma de decisiones.

A diferencia del modelo de Argentina, Brasil presenta una regulación normativa que establece obligaciones entre las entidades estatales para cuyo cumplimiento es necesario contar con datos proporcionados por la evaluación. En este caso, la normativa asigna al Ministerio de Educación la responsabilidad de producir indicadores: de la constitución a la reglamentación del desempeño administrativo, de la distribución de recursos a la definición de indicadores que establecen su distribución, el entramado siempre conduce al IDEB. Dentro de este esquema de regulación, el modelo de Brasil deja poco margen para la negociación porque presumiblemente esta negociación ya se realizó por fuera de ese marco y es allí desde donde estableció sus parámetros. De este modo, sobresale la diferencia con el modelo federal de la Argentina, en el cual la regulación define orientaciones generales y deja las definiciones del uso y de las consecuencias de la evaluación para el plano del consenso en el Consejo Federal y en las provincias.

La Política Nacional de Evaluación y Exámenes de Educación Básica en Brasil determina entre sus objetivos: diagnosticar las condiciones de prestación de la educación básica; verificar la calidad de la educación básica; ofrecer insumos para el seguimiento y mejora de las políticas educativas (Decreto 9432/18). Asimismo, los marcos normativos establecen regulaciones complementarias a la Política Nacional. En su artículo 6, por ejemplo, la Resolución 458 de 2020 del INEP determina entre los objetivos del SAEB las acciones de: evaluar la calidad, equidad y eficiencia de la educación; y subsidiar la elaboración, el seguimiento y la mejora de políticas educativas basadas en la evidencia.

En cuanto a la función de la evaluación orientada a verificar el cumplimiento de metas, la Meta 7 del PNE de Brasil busca promover la calidad de la educación básica en todos los niveles y modalidades de educación y establece metas medidas por el IDEB. Con estas definiciones se encuentran alineados los estados y los municipios, atentos a proporcionar información para el control del cumplimiento de la meta.

Respecto de la coordinación entre actores subnacionales y municipales, Brasil podría caracterizarse como un “modelo de ejecución”, en el sentido de que los criterios de coordinación –tanto presupuestaria como de metas– están estipulados en los marcos normativos y –tal como se señaló anteriormente– se ponen en práctica de acuerdo con los resultados de los índices. El marco constitucional establece que el Fondo de Manutenção e Desenvolvimento da Educação Básica e de Valorização dos Profissionais da Educação (FUNDEB) distribuye un monto de recursos para alcanzar el valor mínimo por alumno estipulado nacionalmente, y de acuerdo a la evolución de indicadores de asistencia y mejora del aprendizaje con reducción de las desigualdades, a ser definidos, en los términos del sistema nacional de evaluación de la educación básica. El SAEB define estos términos y los estados y los municipios alinean sus esfuerzos y sus políticas de mejora en función de esos criterios. Para este tipo de modelo se propone la denominación “de ejecución”, porque los resultados de las evaluaciones ponen en movimiento decisiones, como las de distribución de recursos, que han sido tomadas previamente, mientras que otras decisiones, como aquellas sobre los programas específicos orientados a la mejora, están subordinados a los mismos criterios y mecanismos. En este sentido, es clara la diferencia con Argentina, otro país federal, que emplea la información como insumo para la construcción de metas y criterios de actuación por consenso, en tanto los mecanismos de coordinación están abiertos para ser definidos en un espacio de mediación, que se nutre de la información con posterioridad a su emisión.

Chile, por otra parte, despliega su marco normativo mediante una Ley de Educación que establece que el Ministerio de Educación y la Agencia Nacional de Calidad de la Educación velarán por la evaluación continua del sistema educativo. Desde este marco, asigna a la Agencia la responsabilidad de evaluar los logros de aprendizaje de los alumnos y el desempeño de los establecimientos educacionales sobre la base de estándares indicativos. El corazón del modelo es la existencia de un registro público de

establecimientos, un sistema de fiscalización parametrizado por los resultados de la evaluación de los aprendizajes, entre otros, un sistema de sanciones para los casos de incumplimiento y un organismo que cumple con la función de superintendencia.

La Ley de Aseguramiento de la Calidad (SAC) crea la Agencia de Calidad de la Educación y establece la distribución de tareas entre esta y el ministerio. Así se perfila un esquema de apoyo a la mejora de la calidad, en el cual el Estado asume un rol subsidiario que deja, por un lado, la demanda de apoyo a los establecimientos que se ubican en los niveles medios y altos de desempeño e indica, por el otro, focalizar el apoyo ministerial en aquellos de rendimiento bajo o en los ubicados en sectores geográficos con menor disponibilidad de apoyo técnico⁷.

El marco de referencia se complementa con los planes de aseguramiento de la calidad, que son planes plurianuales de política en distintas áreas. En cada instancia de este diseño normativo, se asumen definiciones estatutarias que construyen vectores directos de acción: se asignan responsabilidades de fiscalización, se establece la obligación de elaborar y revisar periódicamente estándares de desempeño institucional, docente y de aprendizaje de estudiantes (Ministerio de Educación), y se definen sanciones a los sostenedores que no cumplen de manera reiterada con los estándares.

Entre las funciones asignadas a las evaluaciones, sobresalen la función de aseguramiento de la calidad, la de verificación de las metas de aprendizaje, la de responsabilización y la de promoción de la cultura de la evaluación. Según el diseño descrito, el uso de los resultados de las pruebas administradas para la mejora de la calidad es la función de apoyo. Esta se estipula porque es la base que permite la demanda de asistencia por parte de los sostenedores de los establecimientos

7 Las escuelas reciben fondos bajo el compromiso de mejorar los resultados y acuden, con estos recursos, a las organizaciones especializadas en asesoramiento. Estas organizaciones están reguladas por criterios o estándares que pautan las características del servicio.

al Ministerio o a prestadores de servicios, y además porque autoriza la intervención del Ministerio en las escuelas con niveles bajos de desempeño. Si bien el diseño asigna un uso inmediato de los resultados, dado que “dispara” la intervención estatal cuando los desempeños no alcanzan los niveles establecidos por el SAC y es obligación de la Agencia “elaborar y entregar información estratégica a los establecimientos, sus sostenedores y las comunidades educativas a efectos de contribuir a mejorar la calidad de sus procesos y resultados educativos” –tal como se verá en la próxima sección sobre el uso efectivo–, el rol subsidiario asignado al Estado por el diseño institucional desdibuja un poco su intervención y deja planteada la pregunta acerca de qué otros usos para la mejora despliegan las unidades públicas y los servicios de apoyo.

Respecto de la verificación de las metas de aprendizaje, se aprecia que los estándares de aprendizaje forman parte de los que se consideran para la evaluación de los establecimientos⁸. En cuanto a la responsabilización, gran parte del sistema está ordenado por la fiscalización, mientras que los marcos normativos establecen la transparencia de la información. En este sentido, la Ley de Educación establece expresamente que “la información desagregada del conjunto del sistema educativo, incluyendo (...) los resultados académicos deben estar a disposición de los ciudadanos, a nivel de establecimiento, comuna, provincia, región y país” (art. 2°).

En el sistema educativo chileno, la evaluación es la gran reguladora, por lo tanto, un componente y una responsabilidad del SAC es fomentar las capacidades de evaluación en todos los niveles y actores mediante políticas, mecanismos e instrumentos para apoyar a la comunidad educativa y a los establecimientos educacionales en el logro de los estándares de aprendizaje y otros indicadores de calidad educativa (Ley de Aseguramiento de la Calidad).

⁸ Se trata de estándares indicativos de desempeño de los establecimientos, estándares de desempeño docente y estándares de aprendizaje.

En el caso de Ecuador, es posible observar cuatro niveles normativos. En primer lugar, se encuentra el nivel constitucional, que anticipa que existirá una institución autónoma para la evaluación integral. Luego se encuentra la Ley Orgánica de Educación Intercultural (LOEI) y sus reglamentaciones, que establecen las características del Sistema Nacional de Evaluación y sus componentes, y que define la distribución de responsabilidades entre la autoridad educativa y el INEVAL. En este segundo nivel, se establece también la función del Auditor Educativo, función que se ejerce sobre los establecimientos y que culmina con la elaboración de un informe de auditoría que se eleva a las autoridades y que debería emplear como insumo los resultados de las evaluaciones SER. El tercer nivel está compuesto por Acuerdos Ministeriales que antecedieron a la LOEI y que establecieron implementar un Sistema Nacional de Evaluación y un Sistema Nacional de Rendición Social de Cuentas. Este proceso se implementa mediante acuerdos interinstitucionales que, en algunos casos, especifican características de las evaluaciones y también sus usos. En este nivel de normativa, interesa especialmente el Acuerdo 001 de 2017, suscripto entre el Ministerio de Educación, el INEVAL y la Secretaría de Educación Superior, Ciencia, Tecnología e Innovación, ya que el documento acordado entre los organismos determina dos de las funciones de la evaluación Ser Bachiller: la de funcionar como una evaluación estandarizada de grado y la de constituirse como una evaluación para el acceso a la educación superior. Por último, un cuarto nivel contempla las resoluciones del INEVAL que se refieren a cuestiones de logística y de procedimientos. También, existe un Plan Estratégico del INEVAL 2018 y 2021, que se propone varios objetivos, entre los que se cuenta el de incrementar el uso y el acceso a los resultados de las evaluaciones.

En relación con las funciones, los textos normativos ofrecen definiciones generales acerca de la evaluación. Así se pronuncian por la promoción de una evaluación tanto interna como externa que promueva la calidad. Por su parte, la LOEI establece que la finalidad del INEVAL es promover calidad al

sistema educativo y define para el INEVAL las funciones de diagnóstico, monitoreo y evaluación de políticas al señalar que la información resultante de las evaluaciones aplicadas debe servir para la toma de decisiones de política educativa, para el mejoramiento de la calidad y para el seguimiento de metas de corto, mediano y largo plazo.

Entre las funciones mencionadas, la de rendición de cuentas pareciera ser la más destacada dentro de la normativa, a juzgar por su presencia y por su especificidad en los distintos niveles enumerados. Dentro de este esquema, las pruebas Ser Estudiante y Ser Bachiller aplicadas en Ecuador forman parte de un sistema nacional de rendición de cuentas que abarca a diversos componentes. En el plexo normativo, la función de responsabilización de los distintos actores adquiere un peso importante en el marco de un sistema de auditoría para el cual las evaluaciones estandarizadas aplicadas a los estudiantes conformarían un insumo central. Ya en el Acuerdo Ministerial 0450-13 que creaba el Nuevo Modelo de Apoyo y Seguimiento a la Gestión Educativa, se señalaba que el Auditor Educativo debía usar las evaluaciones realizadas por el INEVAL, como uno de los insumos de las evaluaciones externas realizadas en los establecimientos educativos.

En el caso de México, el diseño institucional correspondiente al período 2013-2016 presentaba cuatro niveles normativos: el nivel constitucional; la Ley General de Educación junto con la Ley del Instituto Nacional para la Evaluación de la Educación; el Programa Sectorial de Educación 2013-2018 y la Política Nacional de Evaluación Educativa (PNEE). Dentro de este orden, la Constitución asignaba al INEE, un organismo autónomo y con presupuesto propio, la facultad de coordinar el Sistema Nacional de Evaluación Educativa (SNEE), con la finalidad de garantizar la prestación de servicios educativos de calidad. En tanto, la Ley General de Educación reconocía al INEE la autoridad en materia de evaluación (art. 29). Asimismo, por medio de una ley específica que regulaba al INEE, se establecían los fines de evaluación del Sistema Nacional, entre

los que se contaba la contribución a la mejora de la calidad, al diseño y la implementación de políticas, al control del cumplimiento de metas, a la mejora de la gestión escolar y a fomentar la transparencia y la rendición de cuentas.

Pero las modificaciones establecidas a partir de 2018 reforman el diseño institucional que actualmente se plasma en tres niveles normativos, el nivel constitucional, el de la Ley General de Educación y la Ley en materia de Mejora Continua de la Educación y del Programa Sectorial de Educación 2018-2024. Con este diseño institucional, se crea un Sistema Nacional de Mejora Continua de la Educación, coordinado por un organismo que ocupa el lugar del extinto INEE. La reforma de la Ley General de Educación, por otra parte, establece que se expedirá una ley para regular el Sistema de Mejora Continua de la Educación y que tal ley determinará las funciones de la Comisión Nacional para la Mejora Continua de la Educación, un organismo descentralizado, no sectorizado, con autonomía técnica, operativa, presupuestaria, de decisión y de gestión, que contará con personalidad jurídica y patrimonio propio, y será el encargado de coordinar el Sistema. Asimismo, coordinará el Sistema Integral de Formación, Actualización y Capacitación, el cual será retroalimentado por evaluaciones diagnósticas.

Sin duda, la revisión del diseño institucional de la evaluación en México expone una reorientación en el modelo de evaluación. El repaso por la normativa pareciera indicar que los actuales marcos regulatorios reducen sus alcances estatutarios y son menos precisos en sus prescripciones a los actores. Las normas enuncian las funciones de la evaluación con un menor nivel de especificidad que el diseño previo, también omiten la mención al control del cumplimiento de los objetivos del sistema y las referencias a la rendición de cuentas. En cambio, se mantienen las formulaciones generales para la mejora, aunque ya no se menciona la calidad en la educación, tampoco se observan referencias al uso de los resultados para el diseño de programas específicos, como los relativos

al abandono, a la identificación de poblaciones con rezago en aprendizajes o al diseño de estrategias de desarrollo profesional docente que se encontraban de modo explícito en el diseño anterior.

Es evidente que se reformula el alcance del Sistema Nacional de Evaluación Educativa, el cual al ser reemplazado por el Sistema Nacional de Mejora Continua de la Educación orienta la evaluación hacia la mejora y no explicita la relación entre la evaluación de los aprendizajes y las instancias de planeamiento, si bien se establece que la evaluación diagnóstica retroalimentará al Sistema Integral de Formación, Actualización y Capacitación.

La disolución del INEE y su reemplazo por la Comisión Nacional para la Mejora Continua de la Educación también diluye la emisión de directrices que había sido asignada al primer organismo a propósito del uso de las evaluaciones para el planeamiento de la política educativa y circunscribe las funciones de la Comisión a la realización de estudios e investigaciones especializadas y a evaluaciones diagnósticas y formativas.

Al analizar el caso de Uruguay, es posible observar que el marco normativo que organiza las evaluaciones nacionales y las evaluaciones formativas es aquel que regula la actividad de la DIEE, una Dirección dependiente del área de planeamiento con funciones asesoras del Consejo Directivo Central (CODICEN). La DIEE se constituye como la estructura responsable de las evaluaciones mencionadas y también de las evaluaciones internacionales. La evaluación de los aprendizajes está caracterizada en los marcos normativos como la evaluación del rendimiento educativo de los diferentes subsistemas de la ANEP, con el propósito de devolver a los actores información destinada a la mejora continua del trabajo docente. Este mismo marco establece que la DIEE deberá proporcionar información sustantiva para la planificación, la coordinación, el seguimiento y la evaluación de los

objetivos estratégicos que están bajo la responsabilidad de la ANEP. El rasgo principal que caracteriza a Uruguay es su clara definición por un modelo de evaluación formativa y el uso de las evaluaciones para producir conocimiento acerca de los aprendizajes.

Por otro lado, el INEEd es un organismo creado por ley que tiene como responsabilidad evaluar la calidad de la educación nacional. Con este fin, la ley le señala los siguientes objetivos: evaluar los niveles de educación obligatoria, aportar información que contribuya a garantizar el derecho a una educación de calidad, dar a conocer el grado de cumplimiento de los objetivos y las metas establecidos, favorecer la producción de conocimiento sobre los procesos de evaluación, aportar información acerca de los aprendizajes de los estudiantes, proponer criterios y modalidades de evaluación y asesorar sobre evaluaciones internacionales al Ministerio de Educación y Cultura, y a la ANEP. Cada cuatro años se tiene que presentar un plan estratégico y cada dos años se consolidan informes acerca del estado de la educación en el país.

En relación con las funciones, hasta el momento, el INEEd y la DIEE tienen repartidas las pruebas aplicadas en el país, sin embargo, no está clara la distribución de funciones que uno u otro organismo deben cumplir en relación con la evaluación y pareciera que tiene algunos roles superpuestos. A primera vista, los objetivos del Instituto abarcan propósitos más amplios que los de las evaluaciones gestionadas por la DIEE, como dar a conocer el grado de cumplimiento de las metas, no obstante, el grueso de las funciones continúa orientado por el modelo de evaluación formativa.

Finalmente, en relación con la dimensión estatutaria y su grado de prescripción, en Uruguay la normativa para las evaluaciones es genérica. Esto supone que no prescribe a los actores acciones a desarrollar a partir de los resultados, por lo que es verdaderamente difícil atribuir a los datos obtenidos posibles usos para el planeamiento de políticas.

A modo de síntesis

A partir de lo expuesto es posible destacar algunos rasgos de cada uno de los dispositivos según cómo se presenta la normativa analizada. La pregunta que guía esta síntesis es en qué medida los mandatos regulatorios establecidos en los marcos normativos prevén usos y cómo se relacionan estos mandatos con la toma de decisiones.

Argentina, como se ha señalado, presenta un diseño institucional en el que la evaluación pareciera orientarse predominantemente a la realización de diagnósticos que apoyen las instancias de planificación federal. En algún sentido, podría decirse que presenta un diseño abierto e inacabado, a juzgar por las funciones delegadas al Consejo Nacional de Calidad Educativa y un desacople entre la función argumentativa y la prescriptiva, en cuanto quedan mediadas por la voluntad de consenso de los actores. Los mandatos regulatorios indican alcanzar acuerdos y coordinar políticas, pero no hay implicancias inmediatas para los actores, que se desprendan de los resultados. A diferencia de Brasil y de Chile, países en los que los acuerdos políticos están por fuera de los sistemas de evaluación (en instancias superiores, legislativas o en acuerdos federativos), y donde estos son piezas para monitorear su cumplimiento, en Argentina el diseño sería inverso, y los acuerdos políticos –incluso los federales– parecieran alimentarse de los resultados de la evaluación, en tanto que esta debería proporcionar insumos para la discusión y para la construcción de esos pactos. Las normativas de las dos jurisdicciones subnacionales analizadas, por otra parte, enfatizan el uso de los resultados de evaluación para la realización de diagnósticos y para la mejora educativa.

Brasil presenta una estructura normativa densa y articulada, anclada en el nivel constitucional y centrada en la verificación del grado de cumplimiento de las metas de aprendizaje, consolidando un mecanismo mediante el cual se vinculan actores subnacionales. Las normativas estatales también son densas

y muchas de ellas orientan un uso de evaluaciones con altas consecuencias. Como se ha mencionado, el de Brasil puede caracterizarse como un modelo “de ejecución”, dado que los resultados de las evaluaciones ponen en movimiento decisiones acordadas previamente. El dispositivo normativo, por otra parte, establece mandatos regulatorios de ejecución inmediata, a través de la aplicación de los resultados de las evaluaciones de larga escala y de sus correspondientes índices.

Chile, a partir de un Estado que se define como subsidiario, delimita una distribución de funciones en virtud de la cual la verificación del grado de cumplimiento de metas y la obligación de proporcionar información pública para la responsabilización corresponden al Sistema de Aseguramiento de la Calidad, mientras que las autoridades educativas desempeñan la función de acompañar la mejora de la calidad. En este caso, el mandato más fuerte para la autoridad central pareciera ser el de ejercer la función fiscalizadora, mientras que el mandato de la mejora recaería sobre los establecimientos. En cuanto a las funciones diagnósticas y de monitoreo, estas parecieran distribuir por partes iguales entre los niveles macro y micro de las políticas.

En Ecuador, la normativa pareciera haberse inclinado originalmente por un modelo de responsabilización, sin haber establecido, no obstante, las consecuencias de la falta de cumplimiento de las metas ni un mecanismo de ejecución de tales consecuencias. El diseño del sistema de evaluación fue antecedido por un plebiscito y, en 2008, la Constitución estableció la creación de un instituto autónomo que luego sería el INEVAL, en tanto, por un Acuerdo Ministerial se creó en 2009 un sistema nacional de rendición de cuentas (SER). Las leyes posteriores retomaron estos mandatos, pero no parecen haber especificado prescripciones en términos de responsabilización, por lo tanto, sus usos quedaron acotados, por defecto, al modelo formativo. En efecto, la acción se dirigió hacia una función de implicancias altas, sin un mandato claro para su ejecución. Por otra parte, el dispositivo de pruebas Ser Estudiante y Ser Bachiller divide las funciones para este último nivel, en

evaluaciones con función diagnóstica y con funciones de certificación de finalización de estudios, una decisión que genera obstáculos durante su aplicación efectiva.

México pareciera estar en transición desde un diseño orientado a la evaluación integral con una función de coordinación federal, hacia un diseño diagnóstico al servicio de un modelo formativo. Esta modificación pareciera instar a un desplazamiento de las funciones asignadas a las pruebas estandarizadas, el cual iría desde un uso en el nivel macropolítico y en las entidades federativas—que no tienen un papel claramente asignado en los nuevos marcos—hacia un uso en el nivel del aula, mediante el empleo de pruebas diagnósticas. Al mismo tiempo, se sugiere que el reintegro de las responsabilidades del ex INEE a la Secretaría de Educación tiende a resolverse en una delegación hacia los establecimientos educativos, aunque sin un marco de responsabilización, como ocurre en el caso de Chile, o con efectos en la distribución de recursos, como en el caso de Brasil. La información recabada sobre los recientes marcos normativos no permite caracterizar el papel asignado a las entidades federativas, que no aparecen mencionadas especialmente en relación con las instancias de evaluación.

Finalmente, Uruguay presenta un esquema con dos dispositivos de evaluación, ambos alineados con la evaluación formativa, según la definición general de este país. Las evaluaciones originadas en ambos dispositivos tienen alcances diferentes: mientras que las aplicadas por la ANEP son censales, anuales y voluntarias, las aplicadas por el INEE son muestrales, bianuales y obligatorias. Según las funciones asignadas, los usos de ambos dispositivos son formativos, orientados al diagnóstico, el seguimiento y la mejora de los aprendizajes. En consecuencia, los mandatos regulatorios sobre los actores del sistema educativo están centrados en orientar esta función.

5. Usos efectivos de las evaluaciones a gran escala

En este capítulo, se presentan los usos de la información proporcionada por las evaluaciones estandarizadas de aprendizajes nacionales e internacionales. Se los presenta bajo una clasificación en ocho tipos, de acuerdo con la información emergente de los seis estudios nacionales. Esta taxonomía se basa en la elaborada por Cecilia Galas (Galas, *et al.*, 2020) si bien ha sido ajustada mediante la supresión y el desagregado de algunas categorías, para permitir la aparición de los hallazgos reportados. A continuación, se presentan en una lista los usos predominantes en los seis países, organizados según esa tipología en la medida que los casos presentan un uso relevante en la categoría que se expone. Para profundizar en los detalles, se remite a los informes de cada país.

1. Planificar estrategias en distintos niveles del sistema.
2. Producir conocimiento (identificar problemas educativos, áreas de intervención y asociar factores con los resultados de los aprendizajes).
3. Desarrollar políticas y programas de apoyo, generales y focalizados, para la mejora de los resultados.
4. Orientar la definición del currículo y la elaboración de materiales pedagógicos.
4. Formar y fortalecer la profesionalización de los docentes.
5. Certificar estudios y realizar pruebas de selección para el ingreso a la educación superior.
6. Evaluar y/o incentivar a docentes.
7. Ordenar posiciones según niveles de desempeño.

Usos efectivos reportados de los resultados

5.1. Planificar estrategias

Se incluyen en este tipo de usos aquellos que se destinan al diseño de planes plurianuales cuando establecen prioridades, metas o proporcionan diagnósticos, cuando definen criterios de asignación de recursos a entidades subnacionales y cuando se emplean para evaluar los resultados de políticas generales o focalizadas.

Entre la mayoría de los entrevistados se enuncia como diagnóstico, entendido este tipo de uso como aquel nivel descriptivo del estado de la educación de la población que apoya las aspiraciones generales de la política, tales como incrementar la calidad o reducir brechas de desigualdades. En algunos casos, este uso se especifica cuando los planes plurianuales formulan metas de logro académico basados en resultados de evaluaciones de años anteriores. Si bien tanto las personas entrevistadas en el nivel de responsabilidad política, como las normativas y los documentos orientadores señalan que una función de la evaluación es el monitoreo de los compromisos asumidos en los planes estratégicos, no parecen ser los indicadores de evaluación los parámetros usados de manera sistemática para evaluar el cumplimiento de los planes sectoriales, salvo como se verá, en el caso de Brasil.

Argentina, un país federal al igual que Brasil, subordina el planeamiento federal a la construcción de acuerdos que toman como referencia indicadores de evaluación. Por lo general, los acuerdos federales no establecen planes estratégicos, sino que ordenan a las provincias –como se ha señalado en el capítulo anterior– a formular planes operativos cuya realización es parte de un proceso de coordinación federal sujeto a la construcción de consensos. Esta es una característica de los acuerdos federales, los que, además, frecuentemente necesitan un paso intermedio que debe cumplirse en las provincias. El diseño del federalismo fiscal argentino descansa en una fuerte

concentración de la captación de recursos por parte del Estado nacional y en una significativa descentralización del gasto en la órbita provincial (Morduchowicz, 2010), lo que probablemente explique el hecho de que el Ministerio de Educación nacional cuenta con pocas herramientas para lograr que los estados subnacionales cumplan sus compromisos. Asimismo, el Informe Nacional del país da cuenta de una institucionalización asimétrica de los espacios para la concertación federal en los que están involucradas las áreas de información y evaluación de cada uno de los estados subnacionales, lo que hace aún más difícil la formulación de planes operativos.

Argentina también reporta el uso de las evaluaciones estandarizadas para ampliar diagnósticos y mejorar las hipótesis para la intervención de las políticas. Las referencias sobre este tipo de uso destacan la importancia de la identificación de brechas en subpoblaciones específicas: habitantes de áreas rurales, migrantes, estudiantes según tipo de trayectoria escolar, población según su origen étnico y población estratificada según las condiciones de su contexto y de sus hogares.

En Brasil, la función de planeamiento estratégico está vinculada estrechamente a la participación del sistema de evaluación en el diseño federal. Así, el pacto federativo que establece la distribución de responsabilidades entre la unión, los estados y los municipios está acompañado de un fondo de distribución y compensación de recursos federales hacia entidades subnacionales.

El dispositivo de rendición de cuentas horizontal contribuye a consolidar el uso de las evaluaciones en el planeamiento estratégico. En Brasil, los Tribunales de Cuentas emplean el IDEB para elaborar las matrices de riesgo que definen qué entidades públicas del sector educación serán inspeccionadas. Además, el IDEB es empleado como parámetro de comparación entre el gasto y los resultados alcanzados en los informes de cuentas anuales, con el fin de controlar el cumplimiento de las metas del Plan Nacional de Educación.

En Chile, por definición, el diseño institucional del sistema de evaluación asigna un papel estratégico a los resultados de la evaluación por vía de sus consecuencias sobre la responsabilización de los establecimientos y, consecuentemente, centra en ellos la decisión y conducta de las acciones para las cuales los resultados de la evaluación son un insumo. Este diseño posee como consecuencia un relativamente bajo nivel de uso de la evaluación para el planeamiento sistémico por parte de los niveles centrales y explica, en cambio, una suerte de desplazamiento del uso hacia los niveles operativos y hacia los niveles de gestión escolar, tal como lo reporta el informe de este país. Puede afirmarse entonces que, aunque Chile comparte con Brasil un modelo de evaluación de altas implicancias, una de cuyas funciones centrales es la responsabilización por los resultados, al asignar al Estado una función subsidiaria en relación con la distribución de recursos y el apoyo a los establecimientos, el uso para el planeamiento estratégico adquiere un menor peso que el que posee en Brasil.

En el caso de Ecuador, las pruebas SER se crearon durante la vigencia del Plan Decenal de Educación 2006-2015 y se usaron para afirmar decisiones que ya habían sido tomadas en dicho Plan de manera previa a la primera evaluación. En tanto, luego de 2015, el país no volvió a formular un plan plurianual sectorial. En este país, la información proveniente de las evaluaciones se ha empleado en la planificación territorial bajo el modelo de Análisis de Presencia Institucional en el Territorio (APIT) de la Secretaría Nacional de Planificación y Desarrollo (SENPLADES), que emplea un índice compuesto que mide la complejidad territorial (cantidad de instituciones, docentes y estudiantes) y la información de la prueba Ser Bachiller, que registra el rendimiento estudiantil.

También se observaron en Ecuador, usos destinados a diagnósticos descriptivos con distintos niveles de agregación: provincial, local, institucional, cruzados por grupos poblacionales según variables sociodemográficas –como género, quintil de ingresos, entre otros– o por caracterizaciones del

entorno del estudiante en su hogar, en su barrio, en su comunidad con información proporcionada por los estudios asociados. La fuente más frecuente de los usos mencionados son las pruebas Ser Bachiller, dado su carácter censal.

En este país, se reporta, asimismo, la realización de estudios de impacto de ciertas políticas o programas que han empleado los resultados de las evaluaciones, por ejemplo, la plataforma de aprendizaje asistida en Matemática, estudios de impacto de la aplicación del Bachillerato Internacional en establecimientos públicos y privados, o los resultados de los planes de mejora si bien, según también se reporta, tales estudios no siempre se han difundido. Por último, el Informe Nacional recoge la realización de investigaciones usando modelos de valor agregado que toman en cuenta los factores asociados con el rendimiento de los estudiantes.

En el caso de México, el Programa Sectorial de Educación 2018-2024 recupera los resultados de las evaluaciones de logro como insumo para el diagnóstico de la situación educativa del país reportados por PLANEA 2017, la última evaluación estandarizada aplicada en el país al momento de realizar este estudio. Sin embargo, dado que México modificó en 2018 el diseño institucional de su sistema de evaluación, el país pareciera ingresar en una etapa de revisión del uso de los resultados de la evaluación estandarizada a gran escala.

Durante el período 2012-2018, México promovió una Política Nacional de Evaluación Educativa, con el propósito de impulsar en los estados subnacionales el diseño y desarrollo de Programas Estatales de Evaluación y Mejora Educativa (PEEME). Los PEEME eran parte de un programa institucional que impulsaba, en cada estado, proyectos, objetivos, acciones y metas de evaluación. Si bien no se cuenta con información sobre sus resultados en el ámbito federal, en el Estado de Guanajuato se observa un desarrollo relevante del Programa Sectorial Estatal, con la mención en el diagnóstico de los resultados de las pruebas PLANEA y PISA.

En este país, también se reporta el uso de los resultados de la prueba Enlaces por parte del Programa de Escuelas de Tiempo Completo para identificar si la estrategia ha tenido algún efecto sobre los resultados que obtuvieron los estudiantes, en tanto se emplearon los resultados como insumo para la evaluación del Programa de Becas de Media Superior (PROBEMS).

En Uruguay, cuando las administraciones asumen su mandato presentan ante el parlamento un proyecto de presupuesto y un plan de desarrollo que establecen las líneas de acción básica y la ejecución presupuestaria que, una vez aprobada, autoriza el flujo de fondos para las acciones que han sido comprometidas. El Proyecto de Presupuesto y Plan de Desarrollo Educativo 2020-2024, presentado por la Administración Nacional de Educación Pública (ANEP) en 2020, comprometió en uno de sus objetivos estratégicos, indicadores asociados con metas que fueron construidas sobre la base de los resultados de las evaluaciones estandarizadas Aristas y PISA. Según el plan, las metas deberán ser contrastadas en 2024 con el porcentaje de estudiantes en 3er año, en niveles 1 y 2 de Lectura y de Matemática. Aunque la presentación de planes de desarrollo educativo constituye una práctica regular de las administraciones políticas en Uruguay cuando asumen sus mandatos, esta es la primera vez –según reportan las autoridades entrevistadas– que se emplean resultados de las evaluaciones estandarizadas para plantear metas.

5.2. Producir conocimiento (identificar problemas educativos, áreas de intervención y asociar factores con los resultados de los aprendizajes)

Esta categoría agrupa los usos de los resultados orientados a ampliar y a profundizar el conocimiento disponible sobre las condiciones que influyen en la adquisición de conocimientos de los estudiantes, en el contexto escolar. Este uso supone, por lo general, la integración de los resultados de la evaluación con otra información, como la proveniente de los cuestionarios complementarios aplicados con las evaluaciones estandarizadas, o con información de otras fuentes regulares

con las que cuentan los sistemas educativos. Este empleo de resultados puede impactar en la definición de la agenda, en la formulación de la política o en su implementación, si bien no tiene una ubicación unívoca en tanto depende del aporte de conocimiento que alcance en el marco de una política o un programa en particular.

El uso está asociado con la función formativa de la evaluación a gran escala. Podría caracterizarse también como un uso orientado a incrementar el conocimiento sobre el funcionamiento del sistema educativo, aunque no sea posible sobre la base de la información disponible determinar su uso derivado, ya que para ello sería necesario realizar estudios más detallados.

Argentina, a partir de la información proporcionada por el operativo Aprender aplicado en 2018, registra estudios para identificar aquellas dimensiones institucionales que pudieran constituirse en factores significativos para explicar los resultados alcanzados, como el nivel socioeconómico, el clima y la convivencia escolar. Por otra parte, se reporta la búsqueda de información mediante el análisis de las respuestas del personal directivo, docente y del estudiantado a los cuestionarios complementarios, para obtener datos de interés para la gestión educativa, como la disponibilidad de conectividad, el uso de tecnologías en el aula, la existencia y la valoración de propuestas de formación continua y la recepción de temas vinculados con la implementación del Programa de Educación Sexual Integral (ESI).

En algunos países, se releva la importancia que adquiere el aporte inicial de los resultados de la evaluación a los debates que reorientaron la agenda pedagógica. El SAEB de Brasil, durante la fase de instalación de las evaluaciones, tuvo un papel central en la producción de conocimientos sobre factores que intervienen en el rendimiento académico, lo que produjo efectos en la renovación de la agenda de políticas. También se mencionan varios estudios desarrollados entre

1998 y 2001. Por ejemplo, un estudio de factores asociados, que mostró el peso negativo de la sobriedad en el desempeño estudiantil. Este estudio alertó sobre las consecuencias de que los estudiantes perdieran la continuidad con su grupo de referencia, lo que puso en cuestión la decisión de obligar a la repetición del curso como única solución disponible a los déficits de aprendizaje, reforzó la relevancia de las políticas de reducción de la reprobación e impulsó la búsqueda de medidas pedagógicas alternativas. Otros estudios de factores asociados apoyaron la importancia de las políticas de distribución del libro didáctico o demostraron el peso de la escolaridad de madres y padres en el rendimiento académico. También se debatió acerca de la realización de estudios específicos para comprender por qué el estudiantado, presentaba y aún presenta, en Brasil, dificultades en el aprendizaje de las operaciones matemáticas con decimales.

La producción de conocimiento usando los datos del SAEB es continua y las metodologías se van sofisticando. En esta función, se destaca el uso del Censo Escolar, que releva información individualizada. Se trata de un instrumento muy robusto para asociar la información obtenida con la aportada por los resultados de aprendizaje.

En Uruguay, dado que el enfoque formativo es central en la definición del modelo de evaluación, el empleo de la información de las pruebas aplicadas, con el fin de identificar problemas educativos y factores que expliquen el rendimiento académico constituye un uso reportado frecuentemente. Este empleo es referido tanto en relación con los niveles de decisión política, como con los niveles técnicos y los de gestión escolar. Las áreas usuarias de gestión de políticas de educación básica, por ejemplo, reportan que tanto la evaluación Aristas como las Evaluaciones Nacionales proporcionan conocimiento para contribuir a la toma de decisiones asertivas. Asimismo, varias personas entrevistadas coinciden en destacar que las evaluaciones permiten conocer aquellas dimensiones de la escuela, del aula y del contexto que contribuyen al logro de

los aprendizajes estudiantiles. Esta contribución también es reconocida como aporte de las evaluaciones internacionales PISA y LLECE, reconocimiento que deriva del análisis de los niveles de logro, combinados con la información proveniente de los cuestionarios complementarios.

A diferencia de lo que se advierte en Brasil y en Uruguay, en Argentina, tal como se desprende del Informe Nacional, la función de producción de conocimiento pareciera encontrarse al servicio de resolver el debate público y especializado sobre la noción de calidad ampliada o restringida, y para proporcionar argumentos vinculados con los usos para la rendición de cuentas o para el uso formativo al que la evaluación debería orientarse, según cuáles sean las posiciones de los actores en el debate. Mientras que, en Uruguay, el predominio de este uso está asociado con el peso del modelo formativo en el empleo de las evaluaciones estandarizadas.

En el resto de los países, por distintas circunstancias, los actores no informan usos destacables en estas categorías. No obstante, hay que mencionar el caso de Chile, que pone a disposición de manera amplia la información producida por el Sistema de Medición de la Calidad Educativa (SIMCE), la cual es empleada por los centros de investigación y los *think tanks*. Algunas publicaciones también reportan convocatorias a centros de investigación por parte del Ministerio de Educación, para la realización de estudios con los datos producidos por el SIMCE (Cox y Meckes, 2016).

5.3. Desarrollar políticas y programas de apoyo, generales y focalizados, para la mejora de los resultados

Se clasifican en esta categoría los usos de resultados de las evaluaciones que se constituyen como insumos para el desarrollo de políticas o programas que se orientan a la mejora de los procesos de enseñanza y de aprendizaje en las escuelas. Se trata de una categoría amplia, dado el nivel de enunciación que se recoge de los documentos orientadores

y entre las personas entrevistadas, y porque corresponde a una función general de los sistemas de evaluación que tiene por definición la mejora de los aprendizajes. Para esta categoría se procuran identificar aquellos usos que intentan incidir en –o apoyar a– las prácticas pedagógicas dentro de las escuelas. Pueden referirse a políticas generales o focalizadas. Esto último ocurre cuando la información proporcionada por las evaluaciones estandarizadas se emplea para seleccionar grupos de escuelas con alguna característica particular: por lo general, con bajos niveles de desempeño en algunas áreas de conocimiento, o combinado este rasgo con una o más variables, como nivel socioeconómico, condición geográfica, pertenencia étnica, entre otras. Asimismo, el uso puede corresponder a los tres niveles de gestión: nivel de sistema o estratégico, nivel operativo y nivel escolar. En el nivel estratégico se definen orientaciones generales para las políticas de mejora, mientras que en el nivel operativo se diseñan e implementan programas específicos que forman parte de una política general de mejora de resultados, mediante el apoyo y el acompañamiento de los establecimientos. En el nivel escolar, en cambio, se despliegan usos formativos de evaluación.

Ante la desescalada de la educación presencial durante 2020, debido a las restricciones tomadas para contrarrestar los efectos de la pandemia de COVID-19, varios países reportaron emplear la información de las evaluaciones a gran escala y/o los cuestionarios asociados aplicados en años previos, para identificar poblaciones con condiciones particulares de conectividad o accesibilidad y con niveles de logro que ameritaran la formulación de políticas de atención específicas. En estos casos, las acciones se han clasificado dentro de este grupo de uso de datos.

En Argentina, por ejemplo, se reporta su empleo para renovar y fortalecer las estrategias pedagógicas de programas de apoyo focalizado para la enseñanza de Matemática en el nivel primario, con fuente en el ONE 2013 y en las evaluaciones

Aprender 2016 y 2017. Asimismo, se reporta el uso de los resultados de Aprender 2016 y 2017 como insumo pedagógico para la planificación de las prácticas de enseñanza y de las evaluaciones formativas en los establecimientos, a partir la devolución de la información correspondiente a sus propios desempeños. Por otra parte, el programa Escuelas Faro focalizó un grupo de escuelas primarias y secundarias que concentraban, entre otros indicadores desfavorables, a estudiantes con bajos rendimientos en Lengua y Matemática según la evaluación Aprender 2016. Este programa permitió el diseño de un plan de apoyo a los establecimientos, para el cual se establecieron metas de intervención tomando como referencia los resultados del operativo Aprender 2018.

Además, durante 2020, la información de la evaluación Aprender 2019 se empleó para identificar poblaciones vulnerables en contexto de Covid-19. Con este fin se usaron los cuadernillos complementarios, donde se cruzó la condición socioeconómica de los estudiantes con el nivel de conectividad a Internet en las aulas y se estimaron diagnósticos sobre brechas en el territorio, de acuerdo con los recursos tecnológicos disponibles en las escuelas.

En la Ciudad Autónoma de Buenos Aires, las acciones de mejora se orientan a generar espacios de encuentro que promuevan el trabajo con supervisores y docentes. El trabajo con los resultados está orientado a la comprensión de las dinámicas y los factores que influyen en ellos. Las evaluaciones y la información complementaria se emplean para incrementar el conocimiento de las dinámicas que influyen en los resultados y para concebir alternativas de enseñanza. Por ejemplo, el proyecto “Trayectos de mejora institucional” pone el foco en la dinámica de la organización escolar. Este se inició identificando escuelas primarias que habían sostenido, según las evaluaciones que realiza la Ciudad, mejoras en Lengua y Matemática en años consecutivos. Luego se realizó un estudio de campo en estas escuelas para identificar aquellos factores institucionales que podrían estar influyendo en los resultados y seguidamente

los hallazgos se convirtieron en un bloque de preguntas de los cuestionarios complementarios. A partir de la información recopilada, los equipos técnicos de la Ciudad Autónoma de Buenos Aires trabajan con las escuelas en una línea de mejora institucional progresiva. Las autoridades aclaran que se trata de un proyecto acotado pero que va construyendo un marco de acción. Este proyecto es un ejemplo de programas que integran un uso de los resultados de la evaluación para la producción de conocimiento sobre los factores que influyen en los logros y que desarrollan propuestas de mejora en los establecimientos, en diálogo con las escuelas. La particularidad de este proyecto es el recorte de factores sobre los que interviene, que son específicamente los institucionales.

Por otra parte, la provincia de Córdoba, en Argentina, informa el uso de datos en el marco del Programa Escuelas Faro, a partir de los resultados de la evaluación Aprender, tal como fuera informado para el nivel nacional.

Argentina se constituye como un caso de uso con funciones formativas, que emplea la información de los resultados en acciones de mejora para establecimientos seleccionados, sobre la base de parámetros que incluyen los niveles de logro, entre otros indicadores. No existen índices de aplicación automática y por lo general los resultados proporcionados por las evaluaciones operan como criterios de selección de establecimientos para focalizar programas, es decir que son parte del diagnóstico de las escuelas que forman parte de dichos programas, y se integran a procesos deliberativos para ampliar la comprensión y para la construcción de consensos. Estos usos están enmarcados, por lo general, en programas diseñados nacionalmente y aplicados localmente.

En el caso de Brasil, los programas más extendidos de apoyo a la mejora son aquellos que proporcionan recursos como incentivos contra la presentación de un compromiso de acción. El Programa Dinero Directo en la Escuela (PDDE, por sus siglas en portugués), por ejemplo, ha entregado un incentivo

monetario a las escuelas públicas sobre la base del diseño de intervenciones que comprometan la mejora de los resultados, aunque sus criterios de asignación pueden variar. El Plan de Acciones Articuladas (PAR), por otra parte, es presentado por los gobiernos subnacionales con el objetivo de mejorar los resultados de aprendizaje mediante el acceso a recursos, entre cuyos criterios de asignación puede estar el IDEB, sea para premiar a los mejores o para ayudar a los que tienen peores índices.

Los dos estados subnacionales y el municipio que componen el caso Brasil de este estudio tienen puntos comunes que comparten con el diseño federal y tienen rasgos específicos. Los tres poseen sus propios sistemas de evaluación, emplean índices complejos además del IDEB y orientan su sistema hacia la responsabilización. Una estructura institucional de apoyo a municipios y escuelas complementa las posibilidades y la presión por el uso.

Brasil, en el nivel federal, se caracteriza por la definición estratégica de asignar recursos a establecimientos usando indicadores de evaluación –como parte de índices complejos que incorporan otros indicadores como el flujo de matrícula– y por el diseño de programas específicos en el nivel federal y estadual, mediante los cuales se transfieren esos recursos a la vez que ofrecen un marco de apoyo técnico. Los programas funcionan dentro de un sistema de consecuencias fuertes, apoyado en los resultados de aprendizaje y en el IDEB y, al mismo tiempo, apoyan las decisiones educativas a implementar. Los estados de Ceará y de San Pablo y el municipio de Sobral (Ceará) reproducen este esquema general y agregan esquemas de incentivos a docentes con sus propios sistemas de evaluación. Asimismo, los estados y los municipios observados por el estudio tienen estructuras institucionales de apoyo a los municipios y las escuelas con márgenes de autonomía técnica variables según el caso. El uso en el ciclo de políticas, para estos casos, además de estar establecido como un punto de partida estratégico en tanto el IDEB es el parámetro bajo el cual se asignan recursos,

también es parte del monitoreo mediante el cual se mide el cumplimiento de los compromisos. En cuanto al uso formativo, San Pablo cuenta con evaluaciones bimestrales para orientar la enseñanza y se emplean los resultados para planificaciones durante el inicio del ciclo lectivo siguiente.

En el caso de Chile, los niveles de logro alcanzados en las evaluaciones de las áreas abarcadas por el SIMCE se emplean para describir el desempeño de los establecimientos y para definir acciones a nivel nacional y a nivel local con el fin de brindar apoyo a las escuelas. Aunque el Informe Nacional no reporta el detalle de programas específicos, sí se reporta que los niveles descentralizados poseen la responsabilidad de viabilizar este uso. Asimismo, se informa que existe la modalidad de entrega de resultados por escuela, a partir de cuya información más detallada, se espera que los docentes encaren acciones pedagógicas. Si bien este uso existe como mención general en todos los países y como aspiración discursiva de las funciones de la evaluación, sería necesario aplicar estudios de profundización en el nivel escolar para conocer detalles sobre su uso. Por otra parte, como ya ha sido señalado, Chile recientemente ha desarrollado pruebas progresivas, lo que permite que los establecimientos realicen un seguimiento más detallado de los resultados del estudiantado. Como en el caso de Uruguay, este es un claro ejemplo de desarrollo de un instrumento de evaluación destinado al uso de los niveles escolares.

En Chile, las acciones de mejora están subordinadas al efecto responsabilización de la evaluación con una relativamente alta autonomía de los establecimientos. A diferencia de Brasil, donde los incentivos a las escuelas están mediados por los estados y los municipios en el marco de un mecanismo de asignación de recursos y con estructuras locales de apoyo, en Chile, el Ministerio de Educación no tiene instancias de gestión que asuman la responsabilidad por los resultados de los establecimientos y son las propias instituciones quienes deben hacerlo. Las recientemente creadas unidades administrativas locales (Servicios Locales de Educación) parecieran no

haberse constituido aún como actores relevantes en relación con este punto. En cuanto al uso formativo de la información, los resultados SIMCE son entregados a las escuelas para que estas procedan a su análisis y se han incorporado las evaluaciones progresivas para que los establecimientos las empleen durante el año, aunque sería necesario implementar investigaciones en los establecimientos para conocer sus usos.

Ecuador, por su parte, reporta la creación de planes de mejora en las escuelas y para realizar el diagnóstico de las instituciones educativas seleccionadas para intervenciones focalizadas basadas en los niveles de logro de las evaluaciones Ser Estudiante. En tanto, no se reportan usos específicos de la prueba Ser Bachiller para esta categoría.

México, que ha reorientado su evaluación hacia las pruebas diagnósticas, define todo su dispositivo actual de evaluación en relación con este uso. De hecho, el área responsable pasó a denominarse Comisión Nacional para la Mejora Continua de la Educación (MEJOREDUC), sin embargo, dada su reciente creación y debido a que el nuevo esquema todavía está en proceso de definición, el estudio no ha recogido usos específicos en este sentido para la etapa actual, ni fue posible obtener mayor información sobre las características técnicas.

No obstante, para el período previo a reorganizar su sistema de evaluación, el país reporta como uso de resultados para la mejora de la enseñanza y el aprendizaje, acciones similares a las ya especificadas para otros países. Cuenta con la figura de los Consejos Técnicos Escolares, órganos técnico pedagógicos colegiados que en cada escuela son los responsables de los logros de aprendizaje de las y los estudiantes, y son quienes tuvieron a cargo emplear los resultados PLANEA como insumo para orientar las estrategias pedagógicas y así motivar la reflexión informada de los actores involucrados. En efecto, PLANEA y sus cuadernillos complementarios han sido usados para la reformulación de prácticas pedagógicas.

También se reporta el uso dirigido a orientar refuerzos pedagógicos por medio de la construcción de guías para docentes y estudiantes. En esta línea, se registraron estrategias de intervención específicas para el área de Lenguaje y Comunicación mediante la distribución de cuadernillos dirigidos a profesores y a estudiantes de zonas donde no había acceso a Internet ni a televisión y las evaluaciones reportaban sistemáticamente bajos niveles de logro. Con motivo de la interrupción de la escolaridad debido a la pandemia por COVID-19 durante 2020, se han revisado los bancos de información sobre ítems para realizar un análisis de los contenidos que históricamente han resultado de mayor dificultad y realizar una selección curricular que priorice y apoye la enseñanza de esos contenidos. Las fuentes utilizadas para esta revisión fueron PLANEA, ENLACE y PISA.

En el caso de Uruguay, como se señaló en la descripción del tipo de uso de resultados anterior, el empleo de datos para la mejora en el nivel escolar puede ser definido como el uso pedagógico de la información que proporcionan las evaluaciones para retroalimentar la enseñanza en cada establecimiento educativo. Como se destacará más adelante cuando se analicen los factores que explican el uso de los resultados de las evaluaciones, existe una discusión acerca de si la devolución de los resultados a los establecimientos constituye información suficiente para implementar usos en el aula o si se requiere de un proceso de mediación y construcción específicamente pedagógicas, orientadas por políticas de enseñanza.

Uruguay también ha desarrollado las pruebas Adaptativas, que son pruebas voluntarias que se aplican en el nivel primario. Los docentes las toman de referencia, adaptan sus ítems y las aplican para obtener información sobre el desempeño del alumnado. Aristas en Clase es otro ejemplo de una evaluación referencial destinada al uso de los docentes en el aula.

5.4. Orientar la definición del currículo y la elaboración de materiales pedagógicos

Se clasifican los usos en esta categoría, cuando las decisiones a partir de los resultados de las evaluaciones implican alguna consecuencia en la política curricular, como la redefinición de estándares, la modificación del peso relativo de algún dominio de conocimiento en el currículo o la revisión de los contenidos o de las competencias de algunas áreas. Esta categoría también abarca la producción de recursos pedagógicos, aunque cuando la producción de tales recursos se realiza para acompañar a un programa de mejora específico, se incorpora en la categoría anterior.

En Argentina, en 2018 la resolución del CFE 342 propuso un Marco Nacional para la Mejora del Aprendizaje en Matemática e Indicadores de Progresión de los Aprendizajes Prioritarios de esta área orientados a la formación de competencias. Estos indicadores fueron un intento de acordar un marco federal que avanzara hacia la formulación de estándares, aunque el estudio no mostró si este marco se concretó en políticas curriculares específicas. Asimismo, durante el operativo de Evaluación Aprender 2018, se incorporaron preguntas en los cuadernillos complementarios con el fin de estimar la recepción de temas relativos a la implementación del Programa de Educación Sexual Integral. La elaboración de materiales de apoyo didáctico para el trabajo en el aula es mencionada por las funcionarias entrevistadas, como un uso derivado de los últimos operativos de evaluación ONE 2013, y de Aprender 2016 y 2017. De todos modos, se debe aclarar que el presente estudio no indagó específicamente, ni en este país ni en los otros, acerca de la manera en que los resultados aportan a la elaboración de materiales de apoyo, ni el alcance de la implementación de esos materiales.

Por otra parte, la provincia de Córdoba informa haber realizado revisiones de las capacidades y de las competencias previstas para los egresados de secundaria tomando como

referencia las pruebas PISA. En este caso, no se trata de un uso basado en los resultados de las pruebas, sino de tomar como referencia las capacidades sobre las que la prueba basa su medición. Asimismo, se han tomado como referencias para el desarrollo de prácticas de evaluación formativa en los establecimientos educativos.

En Brasil, la información proporcionada por las evaluaciones y los estudios viabilizados a partir de la creación del sistema, en su configuración actual de 1997, apoyaron el diagnóstico y la decisión de elaborar una Base Nacional Curricular Común (BNCC), en 2018.

Para San Pablo, se informa que los resultados de las evaluaciones se emplean en la elaboración de materiales y documentos del programa en San Pablo Hace Escuela, un programa que acompaña la implementación del currículo oficial con el propósito de alinear el currículo de las escuelas del estado. Este programa prepara material de apoyo para profesores y estudiantes de los últimos años de la educación básica y media (CINE 2 y 3).

En el caso de Chile, se emplea los resultados de la prueba SIMCE para monitorear la implementación del currículo mediante el control del cumplimiento de los estándares. También se reporta el uso de las pruebas internacionales para realizar el monitoreo del logro de competencias en algunos dominios de conocimiento y para el diseño de recursos educativos como los textos escolares. Un estudio de Cox y Meckes (2016) amplía esta información al describir el empleo de evaluaciones internacionales en las políticas curriculares del país. En primer lugar, el efecto de comparación de los resultados de las pruebas internacionales, junto con otros elementos, habrían impulsado la definición de una política y estándares curriculares y, posteriormente, los resultados de CÍVICA (ICCS, por sus siglas en inglés) habrían impulsado la incorporación o la revisión de dominios de conocimiento como Ciencias de la Tierra y de contenidos de Formación ciudadana. También

informan la formulación de competencias del currículum nacional para el nivel de media básica (CINE 2) bajo el modelo de las competencias de PISA para pensamiento matemático y para ciencias.

En Ecuador se informa de manera sucinta que se hicieron ajustes para el currículo a partir de las pruebas SER.

En cuanto a México, MEJOREDU ha realizado revisiones de resultados de las evaluaciones PLANEA con el propósito de reformular el currículo y para realizar una exploración de contenidos en educación básica en las áreas de Matemáticas, Ciencia, Lenguaje y Comunicación, Formación Cívica y Ética, con el propósito de revisar la carga de contenidos. Al igual que lo ocurrido en Chile, México reporta que se han utilizado los marcos de referencia de la prueba CÍVICA (ICCS, por sus siglas en inglés) para revisar planes de estudios. Una informante clave precisó que algunos conceptos particulares del marco referencial de CIVICA 2009 se incluyeron en el programa de estudios en 2011 y que algunos otros conceptos del marco referencial de CÍVICA 2016 ingresaron en el programa de estudios de 2017.

De Uruguay, no se ha recogido información específica sobre este uso de los resultados de evaluación de los aprendizajes.

5.5. Formar y fortalecer la profesionalización de los docentes

Este empleo de resultados se refiere a los usos de las evaluaciones destinados a la mejora de las acciones de formación docente, se trate de la educación inicial o continua.

En Argentina, los indicadores de desempeño en Lengua y Matemática en los niveles obligatorios aportados por los resultados de Aprender 2018 se emplearon como parte del diagnóstico y como argumento de validación ante la necesidad de desarrollar acciones de formación continua que

tuvieron como destinatarios a los docentes en servicio. Si bien es necesario aclarar que las personas entrevistadas señalan limitaciones metodológicas al momento de abrir los resultados para que estos ofrezcan detalles suficientes que permitan definir contenidos de la formación. Por su parte, la Ciudad Autónoma de Buenos Aires, refiere el empleo de los resultados de sus evaluaciones FEPBA, FESBA, y TESBA, a partir de cuyo análisis los especialistas curriculares y el área a cargo de la formación continua de docentes definen orientaciones para la formación continua. La provincia de Córdoba, por su lado, informa haber tomado una decisión estratégica en relación con la enseñanza de la Matemática que consiste en reorientar las prácticas y la formación docente. En consonancia, el Instituto Superior de Estudios Pedagógicos de Córdoba (ISEP) decidió organizar una carrera específica para formar profesores a cargo de la formación en Matemática.

En Brasil, también se menciona la formación como una de las áreas que se beneficia de los resultados de las evaluaciones, mientras que esta decisión se presenta más enfáticamente en los gobiernos subnacionales y, entre los estudiados, se destaca el caso de San Pablo.

En Chile, los funcionarios entrevistados subrayan la dificultad que se les presenta para vincular resultados académicos con desempeño docente, como un argumento para justificar la ausencia de un uso directo entre resultados y formación docente. No obstante, se reporta que para los programas tercerizados de profesionalización docente, se han empleado los resultados de las evaluaciones PISA, con el fin de priorizar áreas o contenidos específicos de ciertas asignaturas, durante la preparación de las bases para los procesos licitatorios con las universidades.

En el caso de Ecuador, una especialista entrevistada señaló que las pruebas SER Estudiante se utilizan para definir la agenda formativa de los docentes en servicio y que además se registra un caso específico aportado por las evaluaciones

internacionales SERCE y TERCE. En efecto, a partir de los resultados de aplicación de estas pruebas, se obtuvo una descripción del ámbito rural y de la brecha existente entre los sectores urbanos y los sectores rurales y, con esta información, se diseñó una estrategia de capacitación dirigida a las escuelas rurales multigrado y unidocentes.

En el caso de México, la evaluación diagnóstica de estudiantes y de docentes cobra gran relevancia en el presente sexenio para la formulación de programas de formación continua; no obstante, dichas evaluaciones todavía se encuentran en proceso de diseño.

5.6. Certificar estudios y realizar pruebas de selección para el ingreso a la educación superior

Este tipo de uso se refiere al empleo de resultados que certifica al estudiantado la finalización de sus estudios al completar la secundaria baja o alta (CINE 2 o CINE 3) y también cuando los resultados cuentan como examen de acceso a la educación superior. Ecuador ha informado que el uso de las pruebas Ser Bachiller para la certificación de estudios y, en algunos períodos, como parte del puntaje de acceso a la educación superior, constituyen usos que han ocasionado varias controversias en el país.

Fuera de las analizadas para este estudio, Brasil y México cuentan con evaluaciones específicas destinadas a la certificación de estudios y de ingreso a la educación superior. En Brasil, por caso, se aplica el Examen Nacional de la Educación Secundaria (ENEM), mientras que México cuenta con el Instrumento para el Diagnóstico de Alumnos de Nuevo Ingreso a la Educación Secundaria (IDANIS), mediante el cual se proporcionan datos a las autoridades educativas de varias entidades del país sobre el desempeño de los aspirantes a ingresar a este nivel educativo. En este país, también se administra el examen de selección para el ingreso a la educación media superior (secundaria alta) y superior, coordinado

por la Comisión Metropolitana de Instituciones Públicas de Educación Media Superior (COMIPEMS), si bien esta evaluación no es de aplicación generalizada, sino que principalmente se emplea en la zona metropolitana de la Ciudad de México y en 22 municipios conurbados del Estado de México.

5.7. Evaluar y/o incentivar a docentes

Aquellos sistemas de evaluación que se orientan por el modelo con consecuencias vinculan o han vinculado los resultados académicos de las evaluaciones estandarizadas de aprendizaje a sistemas de incentivos docentes. Por lo general, se trata de un factor de medición dentro de un conjunto más amplio de indicadores.

En Chile existe el SNED (Sistema Nacional de Evaluación de Desempeño), mediante el cual los establecimientos mejor evaluados se hacen acreedores, por dos años, de la Subvención por Desempeño de Excelencia destinada a otorgar incentivos remuneracionales a los docentes, y de la Subvención por Desempeño de Excelencia destinada a incentivos remuneracionales para los asistentes de la educación. En este país, el índice tiene asociado un 65% del puntaje del SIMCE a nivel de las escuelas. En el caso de Brasil, el estado de San Pablo definió en 2008 una política de bonificación (para varias categorías de funcionarios públicos), para lo cual el cumplimiento de las metas previstas es un componente del bono.

Desde 2001, en Ceará, el Sistema Permanente de Evaluación de la Calidad de la Educación Básica (SPEACE) entrega premios a directores, técnicos y personal de las estructuras regionales de apoyo y supervisión de redes municipales, y también a directores, docentes, personal y al alumnado de las escuelas estatales.

También el municipio de Sobral puede otorgar beneficios extras a los profesores que alcanzan las metas de aprendizaje de sus alumnos y alumnas mediante la evaluación externa.

En este municipio, el Premio Escola Aprender Melhor otorga reconocimientos monetarios al equipo directivo y a los docentes, desde el último año de preescolar hasta noveno grado, que cumplan con las metas estipuladas de rendimiento promedio, rendimiento mínimo de cada clase y proporción de la participación de estudiantes en la evaluación.

México, por otra parte, tuvo asociado al Programa de Carrera Magisterial las pruebas ENLACE, como un mecanismo para recompensar a los profesores y profesoras que participaban del mismo y cuyos estudiantes obtenían buenos resultados. Sin embargo, esto llevó aparejado un conjunto de efectos adversos y puso en duda la validez y confiabilidad de las pruebas. En este sentido, también se planteó un conflicto entre la SEP y las organizaciones docentes, lo que terminó generando un debate público y académico acerca de la posible distorsión de los resultados de los estudiantes en las pruebas nacionales (Backhoff y Contreras, 2014).

5.8. Ordenar posiciones según niveles de desempeño

En Brasil, el ordenamiento de posiciones de los establecimientos según desempeño es una práctica que fue perdiendo espacio debido a la divulgación de resultados, que fuera acompañada de datos sobre los insumos disponibles y de un índice de nivel socioeconómico. El conjunto de esta información ha permitido observar la incorrección en que caen las comparaciones de posiciones, debido a que no se analiza en profundidad el contexto de las escuelas ni la población a la que atienden.

En Chile, el ordenamiento de establecimientos según su desempeño es una función del sistema de evaluación. En México, por otra parte, las autoridades de los estados subnacionales afirman que los *rankings* entre estados, con el fin de visualizar la posición del propio, constituye una inquietud inmediata luego de la publicación de los resultados de PLANEA.

Sin duda, el ordenamiento según desempeño como uso inmediato de los resultados de las evaluaciones internacionales –y de PISA en especial– es un uso que se observa con distintos énfasis en los seis países, sobre todo, para enunciar diagnósticos. En Chile, las áreas productoras de información reconocen que importa la comparación que surge del ordenamiento entre países y, al igual que en Brasil, México y Uruguay, se valoran las series históricas de PISA por su fortaleza comparativa. En Ecuador, por otro lado, se reporta haber empleado los resultados de SERCE y TERCE en términos comparativos con el resto de los países de la región para demostrar una mejora relativa. Asimismo, los estados subnacionales tienden a emplear la comparación para colocarse en una serie comparativa internacional, como es el caso de la Ciudad Autónoma de Buenos Aires, que además reporta haber usado los resultados de la prueba PISA para su plan plurianual de gobierno.

Usos predominantes

Luego de analizar de manera detallada los ocho usos efectivos de resultados de evaluaciones que fueron reportados, es importante formular el interrogante acerca de cuáles son los usos predominantes que se detectan en los seis países que se encuentran bajo estudio.

Como respuesta es posible afirmar que los resultados se utilizan predominantemente en todos los países analizados como insumos para: 1) planeamiento estratégico; 2) producción de conocimiento e identificación de problemas, áreas intervención y asociar factores a los resultados de los aprendizajes, 3) desarrollar políticas y programas de apoyo para la mejora de los resultados, 4) orientar la definición del currículo y la elaboración de materiales pedagógicos y 5) formar y profesionalizar docentes, si bien estos usos se presentan con diferentes énfasis, en tanto en todos los casos se brinda un uso menor en el caso del último punto.

Las aplicaciones enumeradas se evidencian tanto en aquellos países en los que predominan funciones de altas implicancias, como Brasil y Chile, como en aquellos otros en los que predominan las funciones de bajas implicancias, como es el caso de Argentina, Ecuador y Uruguay, o en México, que se encuentra en movimiento desde un sistema con algunos rasgos de altas implicancias hacia uno de bajas implicancias.

Por otra parte, los usos destinados al punto 6) certificación de estudios, se evidencian en el caso de Ecuador, con algunas ediciones de la prueba Ser Bachiller. En tanto, Brasil y México cuentan con pruebas específicas para esta función. Luego, usos vinculados con las altas implicaciones, como lo es el punto 7) evaluar o incentivar a docentes y el punto 8) ordenamiento de posiciones según desempeño, se advierten predominantemente en países como Brasil y Chile, y en estados subnacionales de Brasil, como San Pablo, Ceará y el municipio de Sobral, cuyos modelos se inclinan por tales funciones, si bien –como ya fuera señalado– los resultados se difunden con ponderadores que permiten analizar los resultados en el contexto específico de las escuelas y de la población a la que atienden. La siguiente tabla resume los usos que han sido descritos y analizados para todos los países del estudio.

Tabla 1. Usos de las evaluaciones por país

USOS EFECTIVOS	PAÍS
Planificar estrategias en distintos niveles del sistema.	TODOS
Producir conocimiento (identificar problemas educativos, áreas de intervención y asociar factores con los resultados de los aprendizajes).	TODOS
Desarrollar políticas y programas de apoyo, generales y focalizados, para la mejora de los resultados.	TODOS
Orientar la definición del currículo y la elaboración de materiales pedagógicos.	TODOS
Formar y fortalecer la profesionalización de los docentes.	TODOS
Certificar estudios y realizar pruebas de selección para el ingreso a la educación superior.	Brasil, Ecuador, México
Evaluar y/o incentivar a docentes.	Chile, Brasil, México y entidades subnacionales de Brasil y México.
Ordenar posiciones (escuelas o países) según niveles de desempeño.	Para Evaluaciones Nacionales: Chile, Brasil, México, entidades subnacionales de Brasil y México. Para evaluaciones internacionales: TODOS.

Fuente: elaboración propia sobre la base de los Informes Nacionales.

Es importante destacar que los usos efectivos de los datos adquieren rasgos diferenciales en cada país, debido a la función predominante de la evaluación como modo particular de regulación de los actores en países federales y centralizados, como así también debido al entramado normativo y al tipo de compromisos y funciones que asigna y al grado de definición estatutaria que adquiera su componente normativo.

A continuación, se proponen algunas líneas analíticas que explican los énfasis observados en los distintos países respecto de los usos efectivos que han sido identificados.

En relación con los usos para el planeamiento estratégico, es atinado indicar usos de establecimiento y control de compromisos, usos descriptivos y para la construcción de consensos. Con usos para el establecimiento y control de compromisos, se hace referencia a aquellos usos donde los resultados de la evaluación son parte de un conjunto de metas cuyo cumplimiento o desvío es comprobado por las autoridades educativas. El uso descriptivo, por otro lado, se refiere a aquellos usos donde los resultados de las evaluaciones forman parte de diagnósticos y del componente argumentativo de los planes sectoriales pero que no se plasman en compromisos específicos que puedan medirse en términos de resultados. Finalmente, los usos deliberativos se refieren a ese tipo de utilización donde los resultados de las evaluaciones constituyen insumos informativos para arribar a acuerdos de política entre actores subnacionales o de diferentes niveles de gestión. A continuación, se revisarán los países en relación con aquellos usos predominantes que permiten ubicarlos en alguna categoría de esta clasificación propuesta. Sin embargo, más importante que la ubicación en una categoría, será el hecho de identificar sus variaciones.

En la descripción precedente, Brasil, junto Ceará, San Pablo y el municipio de Sobral, son los únicos casos que podrían clasificarse en el uso para el establecimiento y control de compromisos de acuerdo con los usos efectivamente relevados. En cambio, México ha virado de un sistema de establecimiento de compromisos entre la federación y los estados subnacionales, cuando estaba en vigencia el PEEME (entre 2012 y 2018) hacia un esquema presumiblemente deliberativo, a partir de pruebas diagnósticas que se han propuesto a través de la MEJOREDUC, aunque al momento del estudio el sistema estaba en pleno rediseño. Uruguay, por su parte, declara que por primera vez el plan plurianual elevado al parlamento en

2020 establece metas de logro académico para 2024 y evidentemente se deberá esperar a ese año, para confirmar si el país avanza en dirección al control de los resultados.

Chile podría ser clasificado en usos descriptivos ya que –como fuera señalado–el sistema de responsabilización recae sobre los establecimientos y las acciones que se generan como consecuencias, es decir que no se trata de acciones de macropolítica, sino que corresponden al nivel del establecimiento. Como señala el estudio del caso chileno, la definición estratégica del diseño de evaluación de ese país consiste en que el sistema se autoajusta. Y cuando la decisión política se orienta a reforzar los recursos hacia aquellos establecimientos que no cumplen con los niveles mínimos de resultados, esta función está al servicio de dicho autoajuste. La revisión periódica de los estándares curriculares a que dan lugar los resultados de las evaluaciones, de alguna manera, corresponderían a un nivel estratégico macropolítico, no obstante, se tratará este punto cuando se modelice este uso específico.

En Ecuador, se observa un uso claramente descriptivo. Como se señaló precedentemente, no está en vigencia un plan sectorial si bien se reportan usos efectivos referidos a diagnósticos que describen la situación de las poblaciones en relación con distintos niveles de agregación, como variables sociodemográficas o caracterización conceptual.

Finalmente, Argentina puede constituirse como un país representante del uso para la construcción de consensos entre actores, dado que los planes plurianuales asignan a las 24 jurisdicciones educativas y al Ministerio de Educación nacional la responsabilidad de formular programas y acciones, para los cuales la información proveniente de las evaluaciones constituye parte del repertorio de insumos presentes en las mesas de construcción de los acuerdos.

El uso para la identificación de problemas y la producción de conocimiento se explica por sí mismo, ya que se desprende

de la función formativa de la evaluación que se presenta en Argentina y por los usos en Brasil y en Uruguay.

En cuanto al uso para el desarrollo de políticas y programas de apoyo para la mejora de los resultados en las escuelas, es posible apreciar que el uso efectivo está directamente relacionado con las estructuras de apoyo existentes y con su continuidad. Es decir que depende de la existencia de estructuras y recursos técnicos que medien entre los niveles de decisión y los estados subnacionales, los municipios o las escuelas. Esto se advierte en la Ciudad Autónoma de Buenos Aires en Argentina, en Brasil y en sus estados subnacionales. Por otra parte, en ambos países se observa que el repertorio de programas posee una mecánica y un alcance diferentes, aunque los estudios no han profundizado sistemáticamente en estas diferencias.

Brasil despliega un conjunto de programas dependientes del Fondo Nacional de Desarrollo de la Educación (FNDE) que, como se señaló, tiene en el IDEB parámetros para definir la participación de los estados o de las escuelas en los programas y la asignación de recursos, mientras que en Argentina el Programa Escuela Faro pareciera haber actuado como un programa puntual que se ha servido de los resultados de la evaluación como insumos pero que, sin embargo, no ha tenido continuidad. En el caso de Chile, en virtud de la autonomía otorgada a los establecimientos, son las mismas escuelas las que procuran apoyo entre proveedores de servicios certificados para tal fin. En este esquema, aquellos establecimientos con peor desempeño son destinatarios de mayores prescripciones, asesoría y apoyo por parte del Ministerio de Educación.

En relación con los usos orientados a la definición del currículo y materiales de apoyo pedagógico, lo que se observa es que –como ha sido señalado– este uso efectivo requiere de la existencia de una política curricular bien definida que establezca la revisión periódica del currículo sobre la base de estándares. Esto implica que los usos curriculares poseen

vigencia cuando están al servicio de una política curricular basada en estándares. En segundo lugar, emerge el papel de los marcos referenciales (aunque no de los resultados) de las pruebas internacionales PISA y CÍVICA, que se emplean como modelo para la revisión del currículo.

En cuanto a las políticas de formación y profesionalización docente, se registra un uso para orientar la definición de programas de formación inicial y, sobre todo, para la formación continua, aunque este último uso se topa con la limitación—tal como lo señala un entrevistado en Chile—de atribuir una relación directa entre el resultado de las evaluaciones y la enseñanza o —como señala una entrevistada en Argentina—de no proporcionar la suficiente información que permita definir contenidos de la formación. No obstante, los estudios nacionales reportan usos referenciales o argumentales de los resultados de la evaluación que se destinan al diseño de programas de formación.

En el caso de los últimos dos usos, vinculados con las funciones de implicancias altas, los resultados tienden a formar parte de los sistemas de incentivo salarial y de ordenamiento de posiciones según lo que esos mismos resultados indican. En el caso de Chile, se observó que los establecimientos mejor evaluados por el SNED (el SIMCE participa con un 65% del puntaje) reciben una subvención destinada a otorgar incentivos de remuneración a los docentes. México ha tenido dispositivos de incentivos docentes asociados con los puntajes de la evaluación estandarizada de logros. San Pablo, Ceará y Sobral, en Brasil, también tienen asociado un incentivo salarial a sus evaluaciones. Este uso despierta controversias entre distintos actores y en particular entre los gremios docentes, que señalan una responsabilización unilateral sobre los resultados cuya explicación, en cambio, depende de múltiples variables, difíciles de deslindar, de condiciones de ejercicio profesional desiguales y de la responsabilidad de otros agentes que van mucho más allá de la tarea de los docentes. Por otra parte, se han puesto de manifiesto los riesgos que conllevan

este tipo de usos, como el de la “inflación de resultados”, un fenómeno que ha sido documentado en México (Backhoff y Contreras, 2014).

En cuanto a los usos vinculados con el ordenamiento de posiciones según los desempeños alcanzados, se observa que todos los países registran este uso para establecer una comparación entre países en el caso de las evaluaciones internacionales, si bien no siempre se alcanza consenso en relación con este uso y, en cambio, es frecuentemente objeto de críticas. Por último, los cuadros técnicos de los distintos países coinciden en señalar que las habilidades superiores que PISA evalúa constituyen una referencia de suma utilidad analítica para los currículos nacionales.

6. Factores que explican el uso

El estudio se inició con el supuesto de que existe un conjunto de factores que explican el uso –o no uso– de los resultados de las evaluaciones. Se partió de un esquema general que identifica factores institucionales, técnicos, políticos y comunicacionales, según una sistematización original basada en diversos estudios regionales (Galas *et al.*, 2020).

Por factores institucionales se entienden aquellas características del diseño institucional y de la dinámica organizacional que explican el empleo de los resultados de las evaluaciones aplicadas. En este estudio, fueron identificados el peso de la trayectoria de institucionalización de los sistemas o dispositivos de evaluación y el peso de las modalidades de coordinación, junto con la trama organizacional de la que forman parte.

Como factores técnicos se consideran la adecuación de los usos a las características técnicas de las evaluaciones, dado que no cualquier uso es técnicamente válido para cualquier evaluación; la disponibilidad de recursos humanos y técnicos para gestionar las evaluaciones y sus usos; y el acoplamiento o el desfasaje entre el ciclo de producción de información y el ciclo de uso de los resultados.

Entre los factores políticos, el que se revela como más importante es el juego de presiones que se ejercen entre los distintos actores del sistema educativo para orientar las funciones de la evaluación hacia usos de responsabilización o hacia usos formativos. En este juego de tensiones, que puede entenderse como una pugna por imponer un modelo de gobernabilidad sobre el sistema, se articulan factores ideológicos, disputas por la legitimidad de los usos de los resultados y de sus interpretaciones y, también, por las modalidades de comunicación.

Los factores comunicacionales se leen según su capacidad de construir usuarios y de contribuir a las necesidades de interpretación y de uso de la información por parte de los actores.

Una característica sobresaliente del modelo de factores es su carácter relacional. Esto implica que argumentos asociados con factores técnicos forman parte de discusiones políticas que impugnan o propugnan usos, y que factores institucionales (como precisamente los procesos de institucionalización) son influenciados por factores políticos, entre otras posibles interrelaciones. A continuación, se ofrece un repaso de los factores y se suman algunas de sus posibles articulaciones.

Factores institucionales

De la comparación de los seis países del estudio, emerge el papel explicativo de la trayectoria de institucionalización y de la coordinación de las unidades productoras de información con una red institucional, ya sea en el nivel central, en el nivel interestatal, en el caso de los países federales, y con organismos de otras áreas gubernamentales.

En relación con los factores trayectoria de institucionalización y sistema de relaciones institucionales, sobresale el caso del INEP en Brasil. Desde 1997, este instituto consolida una trayectoria de reconocida competencia técnica y se constituye como la autoridad en materia de información e investigación dentro del sector educativo del país. Al mismo tiempo, sus mediciones son centrales en la producción del IDEB, índice al que está asociado una serie de funciones distributivas y de coordinación. En cuanto a la red de relaciones interinstitucionales, el INEP se encuentra en el eje de un sistema que organiza la coordinación de la Unión con las entidades federales y ha contribuido al alineamiento de los dispositivos de evaluación locales con el SAEB. Asimismo, dado que las responsabilidades del sistema de evaluación estarán vinculadas con la distribución de recursos, los organismos de control del

gasto público forman parte fundamental de esta red de relaciones, que otorgan espesor institucional al sistema, en tanto utilizan los indicadores de resultados educativos como un recurso de análisis de control del gasto. En Brasil, los estados subnacionales poseen sus propios organismos de evaluación con relativa autonomía, aunque se encuentran alineados con el SAEB.

El caso del INEP puede contraponerse al del ex INEE, en México, un organismo autónomo que fue suprimido por la gestión de gobierno que asumió la presidencia en 2017. El Informe Nacional de México reporta entre los años 2012 y 2017 una serie de vaivenes institucionales que exhiben en ese tramo de su trayectoria signos de alteración de la autonomía institucional frente a la SEP. Entre las situaciones conflictivas que mencionan las personas entrevistadas que son exfuncionarias del instituto, se registra la atribución y la práctica del INEE de emitir directrices dirigidas a las autoridades federativas. Probablemente, esas tensiones generadas, en el marco de una conflictividad política debida al uso de los resultados para un esquema de premios y castigos dentro de la carrera docente –algo que ya fuera consignado en capítulos previos y descrito con más detalle en el Informe Nacional–, junto con cierto cruce de atribuciones de un organismo cuyas funciones originales habían sido eminentemente técnicas, habrían generado problemas de coordinación y minado su autoridad frente a las entidades subnacionales y frente a las representaciones docentes, poniendo en duda sus atribuciones institucionales y su autonomía como organismo. Continuidad, estabilidad en la incumbencia institucional, claridad y persistencia de las funciones técnicas parecieran ser requisitos que especifiquen el atributo más general de la trayectoria institucional de todo organismo del ámbito de la educación de importante envergadura. En efecto, cuando estos atributos no están presentes, la información producida por el organismo queda expuesta a embates que ponen en duda su legitimidad.

La Agencia de la Calidad de la Educación de Chile fue creada cuando el sistema de evaluación presentaba una extensa y consolidada trayectoria. El Informe Nacional de este país reporta que las relaciones entre la Agencia y las unidades usuarias del Ministerio de Educación están protocolizadas, tanto para la formulación de las pruebas, como para la difusión de la información. También se expone que esto da fluidez a las relaciones interinstitucionales, si bien no impide la existencia de dificultades en la coordinación, tal como señala el estudio al referirse a la función de acompañamiento de las escuelas, cuando esa coordinación sufre superposiciones de roles, según lo declaran las personas entrevistadas.

En términos de trayectoria de institucionalización y sistema de relaciones interinstitucionales, puede decirse que la Agencia es relativamente reciente en la historia del SIMCE y que su instalación implicó la afirmación del sistema de evaluación chileno. El Ministerio y la Agencia exhiben, asimismo, un alto nivel de coordinación, facilitada por la protocolización que se acaba de mencionar. En la misma línea, se informa que la división de tareas entre la Agencia y el Ministerio habría posibilitado una mayor complejidad en el análisis de los resultados y en la lectura, principalmente para el nivel micro y meso del sistema educativo. Sin embargo, también se remarca que, al mismo tiempo, habría disminuido la capacidad ministerial para actuar a partir de esos resultados. En el caso de Chile, el problema para el uso de las evaluaciones para el planeamiento de políticas no pareciera estar asociado con el diseño de la institucionalidad del sistema de evaluación, sino más bien con el modelo de toma de decisiones políticas que asume el gobierno de las escuelas y que se basa en el supuesto de autonomía y autorregulación de los establecimientos.

En Ecuador, el INEVAL presenta una trayectoria relativamente reciente y los ajustes que algunos informantes entrevistados de áreas usuarias señalan parecieran responder a búsquedas de equilibrio, propias de una fase de instalación. El informe reporta un impulso inicial que pareciera

haber menguado con el cambio de autoridades educativas nacionales de 2017 y debido a la ausencia de directrices estratégicas sobre las cuales afirmar las funciones iniciales de las pruebas de evaluación. Un uso previsto en el diseño normativo pero que finalmente no se concretó consistía en emplear los resultados de la evaluación en los informes de auditoría para cada uno de los establecimientos, es decir que se trataba de una función asociada con la rendición de cuentas. La doble función asignada a las pruebas Ser Bachiller, dado que se comportan como prueba de certificación para la finalización de los estudios secundarios y su uso para el acceso a la universidad, pareciera haber influido de alguna manera en la valoración de la tarea del INEVAL. A lo dicho, se sumaría la falta de continuidad en las publicaciones de los resultados de las pruebas Ser Estudiante, un rasgo al que se volverá cuando se aborden los factores comunicacionales. A propósito de Ecuador, también se subrayan algunas tensiones en la coordinación entre el Instituto y el Ministerio de Educación, ya que desde ese organismo reclaman una mayor participación en la elaboración de las pruebas, postura a la que las autoridades del INEVAL responden con la afirmación de su autonomía. Por último, otra limitación institucional que se señala para el uso de las evaluaciones la constituye la ausencia de focos relevantes en el diseño de la evaluación, como el de la educación intercultural bilingüe, por ejemplo, además de la falta de apertura hacia grupos de poblaciones prioritarias.

En el caso de Uruguay, se percibe una controversia alrededor de la definición de las responsabilidades del INEEEd, que se enmarca en la ya señalada tensión acerca de las funciones asignadas al sistema de evaluación general en el país. El Informe Nacional señala un cambio en la gobernanza del instituto producido por las autoridades que asumieron a finales de 2019, quienes sugieren asimismo la necesidad de que se incorporen evaluaciones censales –en el momento de realización del estudio se cuenta con muestrales trianuales– que informen sobre los resultados de aprendizaje a cada centro

educativo y que anoticien a las familias y a la sociedad sobre la situación general. Con el mismo criterio, se orientan por situar la responsabilidad sobre las evaluaciones internacionales, actualmente bajo la DIEE, en el INEEd, dado su carácter de evaluador externo. Esta controversia de funciones es alimentada por la idea de que existe una superposición de funciones entre el INEEd y la DIEE. En efecto, personas entrevistadas del ámbito gubernamental y de distintas áreas, así como otros actores por fuera del sector gubernamental, coinciden en señalar este solapamiento.

Como se ha indicado en este informe regional, la convivencia de ambas unidades podría leerse como la manifestación institucional acerca de la coexistencia de diferentes posiciones respecto de la opción entre evaluación externa y evaluación conducida por la propia corporación docente. En este sentido, aunque el estudio no ha detectado que esta coexistencia sea un obstáculo para el uso de la información, sí se observa que se trata de un debate institucional y también técnico que posee importantes consecuencias para el uso de los resultados. Por un lado, se pone en discusión la histórica confidencialidad de los resultados, pero, además, se plantea la pertinencia metodológica del uso de qué tipo de pruebas para qué tipo de función. Más adelante se retomará esta discusión cuando se aborden los factores técnicos.

El estudio confirma que la coordinación entre diferentes áreas de gobierno es central para alcanzar un uso efectivo de los resultados de las evaluaciones y se sugieren algunas precisiones que se detallan a continuación.

En principio, importa considerar tres niveles de coordinación. Por un lado, está la coordinación entre las áreas centrales o la coordinación horizontal. En segundo lugar, se encuentra la coordinación entre jurisdicciones nacionales y subnacionales (estadales o municipales) y, en tercer lugar, la coordinación entre macro, meso y micro nivel.

Respecto de la coordinación horizontal, es posible ordenar los países en dos grupos. Por una parte, están Brasil y Chile y, por otra, Argentina, Ecuador y México. En tanto Uruguay constituiría un caso intermedio entre ambas categorías. La existencia de agencias autónomas, un entrelazado normativo cerrado con mandatos estatutarios para los actores institucionales, la formulación de planes de evaluación plurianuales y la existencia de protocolos conforman un esquema de organización de las relaciones interinstitucionales en el cual la coordinación depende más del sistema que de los actores. Esta situación está representada por los casos de Brasil y de Chile⁹.

En el caso de Argentina, no se encuentran elementos específicos para describir la coordinación horizontal para el período que cubre el estudio –es decir, a partir de 2016–, no obstante investigaciones que cubren períodos previos y que han reconstruido procesos de decisiones en el ministerio nacional, informan de la existencia de una cierta desconexión entre las áreas productoras de información y las áreas pedagógicas y, específicamente, con las integradas por los equipos de especialistas en contenidos curriculares. En el caso de Ecuador, que cuenta con un organismo autónomo, las limitaciones que se señalan para el uso se refieren a la ausencia de planes plurianuales y a cierta indefinición en los objetivos, según señalan los actores entrevistados. Como ya se ha señalado en capítulos anteriores, el Plan Decenal (2006-2015) alcanzó un papel fundante en el dispositivo de evaluación ecuatoriano, luego de cuya vigencia no se formuló uno en su reemplazo. El estudio pareciera señalar, por lo tanto, un rasgo que diferenciaría a Ecuador de Brasil y de Chile: la falta de un plan rector que oriente el uso de las evaluaciones y que sintonice a los actores con sus propósitos. Esta afirmación también sería válida para la Argentina.

⁹ No obstante, algunos testimonios recabados en las entrevistas sugieren no desestimar los vínculos personales entre las funcionarias y los funcionarios de los distintos organismos. Es el caso de Brasil, donde una persona entrevistada señala que se debe tomar en cuenta la importancia de los vínculos personales entre el director del INEP y altos funcionarios del Ministerio, ya sea con la figura personal del ministro o con otros tomadores de decisiones.

En México, la modificación del diseño institucional mediante la supresión del INEE y la creación de la unidad de MEJOREDU, el pasaje del dispositivo de pruebas de evaluación externa a pruebas de evaluación diagnóstica aplicadas por los docentes de los establecimientos y la existencia de un marco normativo general, sin precisiones específicas acerca del funcionamiento del nuevo esquema al momento del estudio, es señalado como fuente de ambigüedad por parte de los actores entrevistados.

El informe de Uruguay, por otra parte, reporta claridad en la asignación de funciones a las distintas áreas productoras y usuarias, si bien se señala que existiría cierta desinformación y cierta desarticulación entre ambas unidades productoras, lo que es atribuido a la ya señalada superposición de roles entre ambos organismos de evaluación. Este país, por lo tanto, se diferencia de los anteriores debido a que la articulación horizontal entre las áreas centrales productoras y las usuarias no sería reportada como problemática, aunque sí se percibe como conflictiva la coexistencia de dos áreas productoras.

De los tres países federales que integran el estudio, se tiene información sobre coordinación vertical en los casos de Argentina y de Brasil, si bien ambos países representan modelos diferentes. Como ya se ha señalado en el capítulo sobre los usos previstos, el diseño federal de Brasil determina un sistema institucional de coordinación entre la federación y los estados y entre estos y los municipios en cuyo esquema el IDEB alcanza un papel destacado, en tanto refleja los objetivos respectivos de dos de los principales mecanismos de asignación de recursos federales, el FUNDEB y el PAR, que se constituyen en incentivos para la incorporación de evaluaciones a gran escala en la planificación de políticas públicas en los estados y en los municipios. En el Informe Nacional, se observa que existe una estructura formal de apoyo y seguimiento “capilar”, formada por organismos de fiscalización, y acompañada por organismos que agrupan a los responsables

de la educación de nivel estadual y municipal¹⁰. Además, Brasil, articula las metas nacionales mediante el PNE (que en su meta 7 establece logros y cuenta con la Base Curricular Común) y los planes estaduais y municipales. Finalmente, los dos estados subnacionales bajo estudio y el municipio de Sobral reportan la importancia de contar en sus estructuras institucionales con áreas de apoyo técnico a los municipios y a los establecimientos escolares.

En el caso de Argentina, sus mecanismos de coordinación vertical están diseñados para propiciar la construcción de acuerdos interjurisdiccionales, que también cubran a la política nacional de evaluación. Asimismo, los márgenes jurisdiccionales alcanzan a la política curricular porque—como se ha señalado—esta se orienta hacia las jurisdicciones dentro de un marco general que permite un amplio margen para la definición de contenidos. Aunque no existen estudios que describan la dispersión curricular entre las provincias, en las dos jurisdicciones que integran el caso argentino se argumenta que las evaluaciones nacionales no siempre están alineadas con los currículos respectivos, lo que se esgrime como una limitación para el uso de sus resultados y constituye una de las razones por las cuales estas jurisdicciones implementan sus propias pruebas de evaluación de aprendizajes. Si bien este rasgo podría caracterizarse como un factor técnico, tiene una raíz en el diseño institucional que también lo explica.

En cuanto a la asistencia técnica para el apoyo a las jurisdicciones y a los establecimientos, podría argumentarse que la estructura de programas nacionales y provinciales que vehiculizan el apoyo a las estructuras provinciales y a las escuelas (algunos de los cuáles se han mencionado en el capítulo de los usos efectivos) explica un uso disperso de los resultados, dado que cada programa los incorpora de acuerdo

¹⁰ El Conselho Nacional de Secretários de Educação (CONSED) y la União Nacional dos Dirigentes Municipais de Educação (UNDIME) agrupan a responsables de educación estaduais y municipales, respectivamente, y forman parte de la negociación con el gobierno federal.

con su diseño específico. Dentro del esquema de construcción de acuerdos, los usos de los resultados de la evaluación para el planeamiento de políticas son, en Argentina, fruto del consenso y también un insumo para su construcción y, junto con la amplitud de los márgenes jurisdiccionales, son parte de un modelo federal con amplia autonomía por parte de las jurisdicciones.

Por otra parte, la tercera modalidad de coordinación, es decir, aquella que se establece entre los niveles macro, meso y micro, se puede visualizar específicamente en los países de gobierno centralizado y o en los estados subnacionales o municipales, cuando estos se encuentran a cargo del gobierno de las escuelas. Chile, con su modelo que espera el ajuste de los servicios por parte de los prestadores según cuáles sean los resultados alcanzados en las evaluaciones, reserva un rol subsidiario al Estado, en la planificación de las políticas de mejora hacia los niveles meso y micro.

Luego, los estados subnacionales abarcados por el estudio muestran modelos diferentes de coordinación. Por ejemplo, la Ciudad Autónoma de Buenos Aires destaca la importancia de la continuidad, la regularidad y la previsibilidad en la entrega de resultados y una política sostenida de construcción de confianza entre los equipos técnicos y los establecimientos educativos. Desde este marco, la lectura e interpretación conjunta de los resultados forma parte de una estrategia de construcción de puentes que permite que los resultados de la evaluación no sean vistos como una amenaza o como información a partir de la cual no se desprende una orientación para la acción pedagógica definida, que es lo que ciertamente demandan los establecimientos.

En el caso de los estados subnacionales de Brasil, se dispone de sistemas de incentivos a las escuelas y a los docentes, en el marco del modelo con consecuencias. En tanto se destaca, en este nivel, la continuidad, la regularidad y la previsibilidad en la entrega de resultados junto con la existencia de

estructuras de apoyo a los establecimientos destinadas a orientar acciones pedagógicas alineadas con los resultados.

Factores técnicos

Sobre los factores técnicos, es necesario destacar algunas cuestiones importantes en relación con las características técnicas de las pruebas y los usos. En primer lugar, los estudios nacionales no reportan cuestionamientos importantes a la calidad de las pruebas, sino a la relación de sus características técnicas con un uso determinado. Así, en Ecuador y en México, se mencionan sospechas vinculadas con filtraciones de pruebas, lo que ha minado la confiabilidad de los resultados.

En segundo lugar, emerge claramente la capacidad técnica como factor vinculado con la disponibilidad de recursos humanos para gestionar las evaluaciones y sus usos. Y, en tercer lugar, existe una discusión, en los países federales, acerca de en qué medida los contenidos que son evaluados en las pruebas representan una muestra confiable de las prescripciones curriculares subnacionales. También es un señalamiento reiterado en los países federales el desfase entre los ciclos de aplicación de las evaluaciones y el uso de los resultados, lo que constituye un factor que disminuye las posibilidades de empleo de los datos obtenidos.

Los estudios detectan, en Brasil, Ecuador, México y Uruguay, una controversia en torno a la conveniencia y a las propiedades de uso de las evaluaciones censales y muestrales. La discusión se da en relación con las funciones de la evaluación y también acerca de la pertinencia de los usos de unas u otras; se trata, en efecto, de una discusión política acerca del modelo de implicancias de las evaluaciones, pero también de una discusión técnica sobre la validez de la información y de su uso, bajo qué intención. En la función formativa, las pruebas muestrales no aportan información por escuela y, por lo tanto, los usos no son apropiados a este nivel de gestión, salvo

que se construyan interpretaciones y procesos de trabajo que permitan enriquecer el aporte de los datos. Está comprobado que los directivos y los docentes restan capacidad explicativa a los resultados si estos no hablan de sus propias escuelas. Algunas pruebas censales, como las de certificación, aportan información sobre el desempeño de los individuos, pero no necesariamente son válidas para la evaluación del sistema.

El argumento sobre las insuficiencias de las pruebas muestrales está presente en todos los países, predominantemente en las discusiones alrededor de las pruebas internacionales, especialmente acerca de PISA, pero también en relación con el uso de las pruebas nacionales muestrales. En México, en el estado de Guanajuato, se señaló la insuficiencia de las pruebas PLANEA debido a su carácter muestral y la preferencia por las pruebas censales, para obtener un seguimiento por alumno y por alumna, y para poder informar a las familias, los docentes y los directivos.

Ya se ha señalado el argumento sobre la inadecuación del uso para certificación que se asigna a las pruebas censales Ser Bachiller de Ecuador, cuya validez como prueba para la evaluación del sistema ha sido cuestionada por algunos actores. En este caso, la complejidad técnica es aún mayor y ello fue señalado durante las entrevistas: una evaluación censal para cursos seleccionados que recabe información de resultados por cada alumno alumna de esos cursos no necesariamente sería válida para diagnosticar la situación de una escuela, si la evaluación se aplica a algunos años. En cambio, un uso apropiado implicaría ponerla en relación con un conjunto de otras informaciones y como parte de un proceso de reflexión institucional orientado técnicamente.

Algunos sistemas de evaluación asumen distinciones para diferentes usos. Como ya se ha señalado, Uruguay posee dos organismos de evaluación y dos tipos de pruebas. México distinguía entre PLANEA ELSEN muestral (a cargo del INEE), para evaluar al sistema educativo nacional, y PLANEA ELCE

(a cargo de la SEP), para uso de la comunidad educativa. En Argentina, el uso orientado de los resultados de las evaluaciones censales pareciera estar presente en el caso de la Ciudad Autónoma de Buenos Aires, donde se reportan instancias de trabajo entre equipos técnicos directivos y docentes de escuelas con quienes se realiza un trabajo de interpretación de resultados mediante una lectura situada en y para cada establecimiento.

En cuanto a la capacidad técnica, es notable la trayectoria de construcción y fortalecimiento de capacidades que se describe en los antecedentes históricos de varios países y en los estados subnacionales que forman parte de la investigación. También se destaca la asimetría en las capacidades técnicas de los estados subnacionales en alguno de los países federales, un tema que es explícitamente mencionado en Argentina, así como la necesidad de fortalecer estas capacidades en las instancias de gestión local recientemente creadas, en el caso de Chile.

Si bien se ha consignado la falta de alineamiento curricular entre las pruebas nacionales y los diseños curriculares subnacionales como una consecuencia de las limitaciones de la coordinación vertical, este rasgo también califica como factor técnico, dado que algunos estados subnacionales y actores no gubernamentales cuestionan su validez y la generalización de sus lecturas. Estos cuestionamientos se profundizan en relación con el uso de los resultados de las pruebas PISA, que se aplican en los seis países. En este caso, los informes reportan como favorables la regularidad en su administración, lo que permite la comparabilidad en el tiempo, y su alta legitimidad técnica, que es mencionada como parámetro por los equipos especializados en evaluación, en tanto algunos países subrayan su uso como referencia para el diseño de la política curricular. Por otro lado, entre sus limitaciones, se señala su carácter muestral, junto con la particularidad de que se basan en competencias por contraposición a los currículos nacionales que están orientados a contenidos. Finalmente, algunos

actores consultados incluso llegan a impugnar la validez de la comparación de los resultados obtenidos por las evaluaciones PISA debido a los distintos contextos de aplicación.

Una mirada más detallada que relaciona el perfil del actor y su perspectiva en relación con las evaluaciones internacionales, demuestra que las visiones favorables o matizadas en general son asumidas por actores que se desempeñan en áreas productoras –y, por lo tanto, expertos en evaluación– o por usuarios gubernamentales y no gubernamentales con formación técnica en el tratamiento de la información. Sin duda, se trata de actores clave del sistema, que tienen claridad sobre la validez de los usos de los diferentes tipos de evaluaciones y pueden diferenciar su valoración de los usos públicos y mediáticos distorsionados, a los que a menudo son sometidos los resultados de las evaluaciones internacionales.

El desfase entre los ciclos de producción, difusión y uso de resultados también es señalado como dificultad entre los niveles federales y subnacionales y constituye la razón que esgrimen los estados subnacionales al momento de marcar limitaciones de uso en el planeamiento de actividades con los establecimientos educativos. Son los casos de Ceará y San Pablo, en Brasil, y Guanajuato, en referencia a PLANEA en México. Estos estados en parte justifican la existencia de sus dispositivos apoyándose en la imposibilidad de resolver el problema del descalce temporal. San Pablo, por ejemplo, ha planificado la entrega de los resultados de sus pruebas SARESP y ha diseñado pruebas diagnósticas de aplicación regular durante el ciclo lectivo, con el fin de acompañar el planeamiento de los establecimientos.

En el caso de Ecuador, otra limitación técnica que se señala para el uso de las pruebas es la ausencia de focos relevantes en la evaluación, como el de la educación intercultural bilingüe, y una apertura por grupos de poblaciones prioritarias.

El tipo de pruebas, ya sean censales, muestrales, destinadas a la certificación de estudios, basadas en contenidos o en competencias, o especialmente diseñadas para el uso escolar, definen cada una un repertorio limitado de usos. A esto se suma que los usos “cruzados”, en el sentido de técnicamente no habilitados, tienden a ser visualizados como no apropiados por algunos usuarios y como que restan legitimidad a los resultados. La falta de alineamiento curricular, aunque sería necesario profundizar más en cada caso, también constituye una problemática relevada y es especialmente destacada a nivel subnacional, en los establecimientos y para las pruebas internacionales. Si bien la baja correspondencia entre pruebas y currículum inhabilita algunos usos, no necesariamente descalifica a los resultados obtenidos. La falta de definición de focos analíticos previos o la ausencia de preguntas relevantes que orienten el relevamiento de los factores asociados se constituye como otro factor que desalienta o desaprovecha los usos posibles.

Factores políticos

Los factores políticos que influyen en el uso –o no uso– de la información proporcionada por las evaluaciones se configuran de acuerdo con la manera como los países resuelven internamente las tensiones entre las funciones de la evaluación, según sus implicancias, tal como fuera argumentado en el capítulo sobre trayectorias comparadas de los sistemas de evaluación. Se perciben entonces factores ideológicos que comandan las pujas entre los actores, usualmente entre funcionarios y exfuncionarios gubernamentales cuando hay alternancia en el gobierno, entre gremios o actores que representan posiciones docentes –en particular, ante usos con efectos regulatorios sobre la labor profesional–, entre funcionarios y funcionarias oficiales, y también entre estos y las organizaciones de la sociedad civil, cuando tales agrupaciones tienen posiciones tomadas en el debate sobre los modelos de evaluación y sus usos.

Dado que la evaluación es un instrumento de regulación de la conducta y de la incidencia de los actores educativos –actores subnacionales, burocracia, gremios docentes, establecimientos educativos–, la administración de pruebas se conforma como un componente central de la gobernabilidad del sistema educativo, lo que genera muchas resistencias y lleva a que se desaten disputas políticas en el terreno de la legitimidad de las evaluaciones y sus usos, junto con importantes controversias técnicas.

Los estudios permiten poner en relación los distintos tipos de factores y la manera como se configuran para favorecer o limitar usos de los resultados de las evaluaciones en los procesos de toma de decisiones políticas. Posicionamientos políticos e ideológicos, conflictos entre actores en torno al uso de la evaluación para la gobernabilidad de los sistemas y el impacto mediático son factores que están presentes en todos los países, mientras que lo que varía es la manera de procesarlos. Como consecuencia, lo que se postula en este apartado es que tanto el diseño institucional, la consistencia entre el instrumento de evaluación y el uso de sus resultados, como la solvencia técnica del sistema, son los atributos que permiten comprender la mayor o menor permeabilidad de los usos de los resultados de evaluación ante los factores políticos.

El espacio mediático por sí mismo conforma una caja de resonancia donde los debates, despojados de su complejidad técnica e impregnados a menudo de una lógica reduccionista, adquieren un peso difícil de eludir para las funcionarias y los funcionarios que poseen responsabilidades directas sobre los resultados de las políticas.

Los diseños institucionales de evaluación representan, –como se ha señalado en el capítulo sobre usos previstos– una solución más o menos estable a una tensión entre dos funciones polares de las evaluaciones, como lo son la función con altas y bajas implicancias. Al respecto, se sabe que Brasil,

Chile y Uruguay representan países con resoluciones cuyas reglas absorben las tensiones entre funciones, aunque en Chile se observa un creciente cuestionamiento. En tanto que Argentina y Ecuador presentan situaciones de equilibrio con reglas amplias y poco prescriptivas (sobre todo en el caso argentino) o con reglas cuyo cumplimiento pareciera no haberse consolidado (como es el caso de Ecuador). Por último, México se encuentra en una situación de revisión de su sistema, que no solo no alcanzó a resolver la escalada de tensiones de la última década, sino que, por el contrario, fue desbordado.

Al profundizar más en el tema, la hipótesis que se sostiene en relación con Brasil es que el país procesó los factores políticos precisamente en la arena política. En efecto, el sistema de evaluación no es el terreno de disputas de posiciones políticas e ideológicas, entre otras cuestiones, porque la evaluación se conforma como una herramienta eficiente en la regulación entre la federación, los estados subnacionales y los municipios. Asimismo, si bien no se ha ahondado en su trayectoria histórica, el estudio nacional de Brasil pareciera sugerir que fue el proceso de consolidación del SAEB el que también tramitó cuestionamientos políticos, ampliando la legitimidad de las evaluaciones y alineando a los sistemas de evaluación subnacionales con la política nacional. De hecho, el estudio no reconstruye los procesos locales que llevaron a varios estados a incorporar incentivos salariales atados a los resultados de las evaluaciones, los que posiblemente hayan generado en su momento debates y conflictos con la representación docente, cuya eventual incidencia en el presente no ha sido reportada. En cambio, el estudio sí registra debates: algunas voces afirman que un sistema centrado en índices con consecuencias fuertes, favorece los *rankings* y los efectos perversos asociados con ellos, y que descuida su función básica que es la de producir información relevante para el cumplimiento del derecho a la educación de calidad de todos los sectores sociales. También emergen posiciones que señalan la necesidad de avanzar en orientar el sistema de evaluación para ofrecer devoluciones más potentes a los

establecimientos y a los docentes, junto con voces críticas que señalan que el peso del IDEB tiende a limitar excesivamente la cultura de la evaluación y la noción de calidad a los resultados de dicho índice hasta reducir las posibilidades del uso formativo.

Chile representa una modalidad de procesamiento de tensiones a través de un diseño institucional que le ha permitido eludir vetos hasta hace poco tiempo, incorporando modificaciones progresivas y manteniendo las funciones originales. Su uso para el planeamiento estratégico pareciera limitado por el papel subsidiario que asume el Estado en el gobierno de los establecimientos educativos; no obstante, tratándose de un modelo con consecuencias, existen usos de los resultados que están inscriptos en su diseño y que son de ejecución directa. En los últimos años, sin embargo, el consenso entre las fuerzas políticas que mantuvo –incluso frente a crecientes voces críticas– el diseño institucional descrito pareciera haber sido desbordado por el estallido social de finales de 2019 y no parece desatinado suponer que se pondrá en revisión con motivo de la reforma constitucional, si bien no es posible prever la profundidad de dicha revisión.

El gobierno del sistema educativo se caracteriza, en Uruguay, por la coexistencia de un Ministerio de Educación y de un ente autónomo, la ANEP, conducido por un Consejo conformado por la representación de la administración del sistema escolar, lo que le confiere una considerable independencia del Ministerio. Ambos órganos funcionan, en algún sentido, como gobiernos paralelos y esta característica pareciera reproducirse en la coexistencia de los dos subsistemas de evaluación.

En Ecuador y Argentina, a diferencia de los casos anteriores, los usos de los resultados de las evaluaciones parecieran ser menos claros. Argentina representa una modalidad de equilibrio en la que se mantiene un debate abierto sobre las funciones de la evaluación, no prescribe usos específicos, aunque excluye modalidades de uso propias del modelo de

altas implicancias, y emplea la información como insumo para diagnósticos, para nutrir debates sobre el derecho a la educación, para componer indicadores de focalización de establecimientos en programas específicos y para usos formativos en los dos estados subnacionales analizados. No obstante, no se alcanza a observar un impacto sobre el planeamiento integral de las políticas educativas.

Ecuador, por otra parte, habría forjado su diseño institucional al calor de una reforma constitucional y de un plan decenal que ordenó un conjunto de funciones del sistema educativo y asignó funciones a la evaluación. Pero el recambio de la administración política pareciera haber debilitado el impulso inicial, ya que el país no formuló un plan plurianual que oriente, entre otras cuestiones, los propósitos de las evaluaciones y el sistema pasó a enfatizar, alternativamente, la evaluación docente, el control del flujo matricular de educación media hacia la educación superior (a través de la prueba Ser Bachiller), en tanto pareciera no haber tenido una política definida para la comunicación de los resultados de Ser Estudiante. Estos rasgos, algunos de los cuales no son privativos de Ecuador, podrían considerarse como indicadores de que los factores políticos han impactado en los propósitos de las evaluaciones circunscribiendo los usos de los resultados a funciones diagnósticas o descriptivas, a componer algunos índices para la planificación territorial y para la certificación de la finalización de los estudios secundarios.

Una comparación entre Ecuador y Argentina puede resultar interesante, porque aunque Argentina representa un modelo de uso formativo de la evaluación y Ecuador asume entre sus definiciones iniciales la responsabilización por los resultados, ninguna de las dos funciones de estos países están desplegadas plenamente cuando se describen los usos efectivos. La función de responsabilización está presente, en el caso de Ecuador, en los marcos normativos, sin embargo, no se recogen entre las personas entrevistadas expresiones de uso concreto en ese sentido. Mientras que las menciones a la función formativa, en el caso argentino, parecieran ser solo

formales, al menos en el nivel federal, aunque exhiben usos concretos en los dos estados subnacionales analizados.

También los aspectos técnicos, que fueran enunciados precedentemente, se combinan con los factores políticos. En este contexto, pueden enumerarse los argumentos “reduccionistas” que, ya sea porque consideran que las consecuencias, las decisiones o las generalizaciones asumidas a partir del uso de los resultados muestrales, ya sea por su alcance, o basados en dominios de conocimiento limitados en relación con lo que se espera de una formación integral, generan impacto. A todo esto, pueden sumarse incluso los argumentos que esgrimen los efectos perversos ocasionados por los *rankings* o los incentivos. En definitiva, todas estas ideas pueden, de acuerdo con la radicalidad del escenario de conflicto y con la manera como los diseños institucionales los procesan, adquirir el suficiente peso como para vetar usos o para canalizarlos hacia transformaciones en el diseño y la administración de las pruebas.

Por último, emerge con cierta sistematicidad un factor político en relación con el uso de las pruebas PISA. Estados subnacionales argumentan la utilidad de los *rankings* para comparar su posición relativa en el concierto de los demás estados subnacionales y, adicionalmente, como una referencia externa que los sitúe en una serie más amplia que les permita, de modo eventual, diferenciarse del contexto nacional. Si bien este juego comparativo puede ser atribuido al efecto de la internacionalización de las políticas educativas y de la globalización, en los estudios nacionales existen indicios acerca de que la utilización de datos se acentúa cuando las fuerzas políticas locales a cargo de la administración de los estados forman parte de la competencia política en el nivel nacional.

Como síntesis, es posible sostener que no es la existencia o la ausencia de factores políticos lo que explica los usos o las interferencias y desplazamientos en esos usos, sino la manera como los diseños institucionales procesan estos conflictos y cómo los canalizan o les otorgan estabilidad.

Factores comunicacionales

Como han registrado los estudios sobre la difusión de resultados desde los primeros operativos de evaluación estandarizada en la región (Cueto, 2005; Ravela, 2001), los casos nacionales recogen la preocupación de las estrategias de comunicación por alcanzar a una multiplicidad de audiencias: decisores, niveles operativos, escuelas, familias, organizaciones no gubernamentales; en tanto registran la necesidad de adecuar la información a las características de esta diversidad. Los seis estudios, sin embargo, ponen de manifiesto que la problemática involucrada desborda la noción de “audiencia”, si se asume que se trata de usuarios cuyo carácter está definido, en parte, por la propia política de uso de la evaluación y de sus resultados.

En conjunto, los seis estudios dan cuenta de los avances regionales en materia de apertura y diversificación de las políticas de difusión, aunque se debe remarcar que estos avances no son homogéneos y que, en algunos países, se observan limitaciones tanto en la continuidad como en la integralidad para la difusión de los datos. Parafraseando a un entrevistado, no existe evaluación exitosa sin devolución.

Las personas entrevistadas en la mayoría de los estudios y las experiencias de comunicación y difusión que pueden considerarse favorables para propiciar el uso de los datos de las evaluaciones convergen en poner de manifiesto la conveniencia de desarrollar políticas de difusión y comunicación continuas, previsibles, que periódicamente incorporen correcciones, actualizaciones y ampliaciones, como parte de una estrategia efectiva para consolidar la valoración de los resultados y construir su legitimidad. Que el conjunto de la información sea concebido como un corpus compuesto por informes nacionales, subnacionales, y sus distintos niveles de agregación, la exhibición de los microdatos en plataformas interactivas con los resultados de las pruebas y de los cuestionarios complementarios –en series históricas–, junto con la confección de manuales y explicaciones metodológicas

e informes de procesamientos especiales, contribuye en su totalidad a expandir la cultura del uso de la información.

En Argentina, Chile y Brasil, funcionarios y funcionarias gubernamentales y usuarios y usuarias no gubernamentales señalan la conveniencia de realizar convocatorias públicas a universidades y centros de investigación, para desarrollar estudios a partir de los resultados de las evaluaciones, con el doble propósito de optimizar capacidades de procesamiento y de análisis y para dinamizar el debate público.

En relación con ciertas particularidades respecto de la difusión, en Ecuador se reporta que desde 2008, con la creación del sistema SER, los esfuerzos en materia de comunicación se fortalecieron, en particular, los referidos a la prueba Ser Bachiller, debido a su uso para fines de graduación, mientras que la difusión de las pruebas Ser Estudiante tuvo algunas inconstancias. También se señala que, aunque están disponibles en la página del INEVAL las bases de datos de los resultados y de las respuestas a los cuestionarios de factores asociados, tanto de las pruebas Ser Estudiante como de Ser Bachiller, desde 2013 y hasta 2019, los informes de esta última evaluación no se presentan con los resultados cruzados con los factores asociados. Esto también pareciera explicarse debido a que quizá se haya privilegiado la función de certificación y de prueba de selección en el caso de las pruebas Ser Bachiller, por sobre la función diagnóstica u otras. En este caso, la comunicación representada constituye un indicador acerca del privilegio otorgado a una de las funciones asignadas a la evaluación.

En Brasil, se reconocen los avances en la estructuración de los datos con un alto nivel de especificación, mientras que, al mismo tiempo, las estrategias de comunicación de los análisis sobre los factores asociados con el aprendizaje todavía no se habrían desarrollado de manera equivalente. Esta observación es parte de la discusión –planteada en Brasil y también en otros países, por los actores de la sociedad civil– sobre aquellas modalidades de difusión de resultados basadas

mayormente en promedios de rendimiento, que ciertamente ocultan desigualdades educativas, al tiempo que se subraya de qué manera la falta de contextualización de los resultados dificulta la formulación de políticas eficaces.

Específicamente puede afirmarse que, respecto del acceso a la información por parte de los organismos no gubernamentales, no se registran mayores restricciones, con excepción de lo que ya fuera mencionado. En los casos de Brasil, Chile y Uruguay, se recogen testimonios acerca de una alta fluidez en el acceso. En Chile, se informa la existencia de un protocolo de pedido de información que resulta funcional y que cumple tiempos razonables de respuesta. Es probable, no obstante, si bien los estudios no son concluyentes, que esta mecánica sea más fluida entre las entidades que se especializan en el tratamiento de datos. Un entrevistado de pertenencia gremial, señala que en Chile no reciben información suficiente o que se enteran de datos de índole educativa por los medios de comunicación, mientras que, en Argentina, algunos usuarios no gubernamentales señalan la importancia de anticipar el tipo de información de resultados que estará disponible, al tiempo que recomiendan difundir en el momento del diseño de los operativos la información acerca de qué y cómo va a ser difundido. Del mismo modo, en Brasil, los estados subnacionales y los medios reclaman una anticipación análoga.

El conjunto de los informes pone de manifiesto la importancia de fortalecer la capacidad de comunicación y de procesamiento de los resultados en los estados subnacionales y en las unidades de gestión local en el caso de los países centralizados. En Brasil y en los estados subnacionales que formaron parte del presente estudio, se observa que el uso de los resultados adquiere fortaleza cuando se emplea en plataformas digitales. Así, los estados subnacionales de este país, apoyados en organizaciones del tercer sector, logran una convergencia entre la devolución de resultados mediante las plataformas, las etapas del calendario de planeamiento y

el proceso pedagógico. En definitiva, se trata de una manera de ensamblar la oportunidad de acceso a la información, con la información customizada y con la tarea pedagógica.

También en México se registró por parte de los estados subnacionales el uso de plataformas diseñadas por organizaciones de la sociedad civil, que formularon reportes por escuela para las evaluaciones PLANEA, aunque –como se ha indicado para esta experiencia– este empleo de resultados habría estado vinculado mayormente con la preocupación de los establecimientos y del Estado (que ha sido su promotor) por mejorar sus posiciones en los *rankings*.

De manera inversa, en Chile se señala que los sectores locales reciben información de distintas fuentes, pero disgregada, fragmentada, y que necesitan incrementar sus capacidades de procesamiento para integrarlas, interpretarlas y para generar un uso adecuado de los datos, con el propósito de “que no abrume y ayude a tomar decisiones”. La inversión de recursos y conocimiento para la construcción de plataformas adecuadas a las necesidades de procesamiento y de uso, tanto de los niveles de decisión local como escolar, pareciera constituir una opción que promueva el uso, si bien la cuestión trasciende la mera problemática comunicacional y se integra al tópico de los factores técnicos.

La problemática de los medios es, entre los factores comunicacionales, particularmente compleja. Por un lado, se constata que los medios, junto con las organizaciones de incidencia, conforman un factor que contribuye a generar un debate público informado sobre la base de los resultados de las evaluaciones y que, como consecuencia, este debate incide en el planeamiento de políticas, lo que también ha sido recogido en el inventario de los factores políticos. Al mismo tiempo, la modalidad de tratamiento sensacionalista en algunos medios suma un argumento para que los gobiernos ejerzan mecanismos de control y para que repliegan la difusión de resultados, lo que retroalimenta –a la vez– el debate público. Este círculo

vicioso pareciera moderarse si se toman en cuenta atributos como la previsibilidad, la continuidad en la difusión de resultados y mediante la presencia de pluralidad de actores en el intercambio público de argumentos e ideas.

Asimismo, y como ha sido observado profusamente en este informe regional, el modelo de las funciones asignadas a la evaluación en gran escala se vuelve relevante a la hora de comprender el papel que desempeñan los factores mediáticos. Aunque el modelo de evaluación, según sea con altas o bajas implicancias, cuando orienta la política de comunicación, influye en los modos de tratamiento mediático de los resultados, también es cierto que existe una relativa autonomía en el debate social respecto de la definición funcional de la evaluación, autonomía que es acentuada por la tendencia de los medios a leer los resultados de manera simplificada. En Chile, donde se registra un incremento del debate público durante los últimos diez años, se señala que la complejidad técnica, acompañada del uso político, complica el debate cuando se difunden interpretaciones a partir de datos absolutos y sin que se comprenda la metodología de análisis de los resultados. En México y en algunos de sus estados subnacionales, el uso de los resultados de PLANEA en *rankings*; en Brasil, el peso y el papel del IDEB, junto con las consecuencias que emergen de los resultados de las pruebas Ser Bachiller aplicadas para el egreso del alumnado del nivel secundario, en Ecuador, proporcionan argumentos suficientes para comprender en qué consiste una difusión de alto impacto mediático.

Sin duda, existe un nivel en la comunicación que sirve para un uso de diagnóstico escolar, mediante el cual un directivo se informa de los resultados de los estudiantes de su escuela y, si bien ya se ha señalado la dificultad técnica que implica realizar esta lectura de manera lineal, aún es necesario agregar que, superada esta limitación, no necesariamente se derivan del diagnóstico realizado orientaciones para guiar la acción pedagógica. Comprender un ejemplo como el planteado implica

que un uso comunicacional que suponga que la ubicación de un establecimiento en un nivel de logro comparado constituye información suficiente para el actuar pedagógico del equipo directivo se basa en dos supuestos que son reduccionistas. Por un lado, se infiere que hay una continuidad automática entre la información y la acción, es decir que solo por conocer el problema se alcanza una solución y, por otro lado, se presupone un gran nivel de autonomía en el rol del director o del docente para operar pedagógicamente con esa información.

Si desde la política se asume que el diagnóstico no indica por sí mismo una resolución pedagógica, sino que entre ambas instancias media un proceso de construcción de soluciones o de adaptación de las soluciones pedagógicas existentes, la cuestión puede formularse de la siguiente manera: ¿en qué medida la presentación de resultados está ensamblada con políticas pedagógicas? Si se trata de contextos de autonomía escolar y docente, como el del modelo chileno, la cuestión incluso se desplaza hacia interrogar cuáles son las condiciones que las políticas se proponen generar –o suponen disponibles– en los contextos locales.

Como consecuencia, es posible apreciar que el problema para el uso de resultados en el nivel de escuela no es tanto –o no lo es únicamente– un problema comunicacional, sino que se debe considerar en qué medida, como parte del diseño de las políticas de mejora (para tomar uno de los usos efectivos mencionados en el capítulo correspondiente), se ha contemplado la función de producción de política de enseñanza (Terigi, 2004).

Este último planteo está relacionado con la comunicación para la retroalimentación pedagógica, lo que vuelve sobre las reflexiones iniciales de este apartado acerca de las limitaciones que conlleva reducir la devolución de los resultados de las evaluaciones a un problema meramente comunicacional. En Brasil, se señala que la formulación de una retroalimentación pedagógica constituye una agenda pendiente, más allá

de los esfuerzos realizados para difundir algunos ítems de las pruebas y para relacionar estos ítems y sus respuestas con lo que efectivamente pueden hacer los estudiantes, en tanto se procura transmitir estas interpretaciones al lenguaje del profesorado. Todavía es necesario destacar la diferencia, que emerge con cierta claridad en algunos de los casos nacionales, entre lo que se concibe como una “estrategia de comunicación de resultados” y lo que debe significar una “estrategia de trabajo con los resultados”, como herramientas –ambas– para la mejora. Esta cuestión pareciera hacerse evidente con la extensión de la preocupación de los sistemas de evaluación sobre la producción y definición de estrategias para uso pedagógico. En jurisdicciones provinciales de Argentina, por ejemplo, se han descrito usos efectivos a través de programas que integran, de alguna manera, la estrategia de comunicación con la interpretación y la producción de sentido de la información, de una manera situada. En palabras de un funcionario del municipio de Sobral, en Brasil, de lo que se trata es de “pasar de evaluación de los aprendizajes a una evaluación para los aprendizajes”. Esta es la manera, por otra parte, como define Uruguay sus evaluaciones formativas. También de manera similar, en el estado de San Pablo, se considera que todavía es necesario mejorar la retroalimentación en relación con las evaluaciones formativas y de diagnóstico, lo cual tiene sentido, dado que en ese país la evaluación sumativa está inscrita en el modelo con altas implicancias que caracteriza al uso de las evaluaciones tanto en el nivel federal como estadual. Como se recordará, San Pablo, al igual que otros estados de Brasil, posee un sistema de incentivos salariales docentes.

A partir de lo expuesto en los últimos párrafos, puede ser útil distinguir dos niveles de análisis que faciliten el hecho de incluir a los medios como factor comunicacional en el uso de los resultados. Por un lado, está el nivel de una política de comunicación que incluya a los medios como usuarios de la información y, simultáneamente, como formadores de opinión pública y actores que participan en el debate. Por otro lado, se deberá tener en cuenta que diseñar una estrategia de

comunicación en el marco de un programa de política determinado requiere tener una política de uso y no solo una estrategia de comunicación, ya que ambas cosas no son lo mismo. Si el primer nivel es el correspondiente al de los medios, el segundo nivel forma parte de las maneras en que la política y sus programas definen el uso para los diferentes tipos de actores.

Finalmente, si bien se sabe que los estudios no entregan conclusiones que permitan realizar una correlación entre el uso de resultados y los factores comunicacionales, sí contribuyen a señalar cuáles son las preocupaciones y las tendencias en la comunicación de resultados que están presentes en cada uno de los casos y ponen de manifiesto los límites de la categoría “comunicación” para concebir mejores modalidades de acceso a la información, por parte de distintos tipos de usuarios. Al respecto, este apartado sugiere la conveniencia de pensar en la comunicación como una construcción de los usuarios, en lugar de considerarla un medio para alcanzar audiencias. Se trata de comprender el papel de los medios como actores y de no reducir las necesidades de interpretación y uso de la información a un problema de transmisión y exposición de datos. Como síntesis, se subraya que será sumamente importante contextualizar la estrategia de comunicación y situarla en una política de uso de la información que vaya más allá de una mera estrategia de difusión de resultados.

7. Conclusiones

Los resultados de las evaluaciones contribuyen con diagnósticos que describen el comportamiento del sistema educativo en función de niveles de logros académicos por sectores de población, identifican brechas en el cumplimiento del derecho a la educación o fortalecen hipótesis de intervención sobre estos u otros problemas. También constituyen la base para el establecimiento de metas de mediano plazo, aunque son menos frecuentes los reportes sobre el monitoreo de tales metas: la evidencia recabada pareciera indicar que los datos de las evaluaciones participan del ciclo completo de política, siempre que estén previstas consecuencias a partir de los resultados y cuando existen los mecanismos que establecen el proceder a partir de esas consecuencias.

Los usos destinados a la coordinación entre actores nacionales y subnacionales han sido reportados con distintos énfasis para los países federales. A la vez, la coordinación pareciera ser una condición cuya presencia o ausencia explica el uso –o no uso– de los resultados de las evaluaciones para la articulación de política. La fluidez de esta coordinación también sería explicada a partir del papel que la evidencia –de la cual los datos de la evaluación son una parte– cumple en el dispositivo de coordinación.

Existen usos, quizá más extendidos y también más fragmentados, que se refieren al empleo de los resultados de las evaluaciones a gran escala como insumos para el diseño y para la implementación de programas de mejora de logros, sobre dominios específicos de conocimiento, como Lenguaje (según sus denominaciones en cada país) o de Matemática, que figuran entre los más frecuentados. De menor intensidad y con menor presencia en los reportes, se observa el uso de los datos en programas de formación docente continua, para

enfocar la agenda de actualización profesional o para diseñar convocatorias a universidades o centros de formación que prestan servicios dirigidos a la profesionalización docente. También se verifica su uso en la implementación del programa de formación, cuanto pueden constituirse en insumos para la preparación de recursos pedagógicos, como libros u otras herramientas para la enseñanza. Una lógica similar se aplica a programas focalizados que seleccionan grupos de establecimientos que exhiben indicadores de logro por debajo de un nivel aceptable, combinados con otros indicadores que también señalan condiciones desfavorables de otras variables de la escolarización. En estos casos, los datos proporcionados por las evaluaciones se integran a un conjunto de evidencias e informaciones más o menos estructuradas sobre problemas que ya están en la agenda pedagógica, refuerzan diagnósticos previos y además suman argumentos a decisiones ya tomadas o en proceso de construcción.

En este sentido, la presencia de los datos proporcionados por las evaluaciones para la formulación de políticas pareciera localizarse sobre todo en el desarrollo de programas específicos e, incluso, es posible suponer que, en ciertos casos, este uso se explique a partir de exigencias de la formulación, cuando se trata de programas que poseen financiamiento multilateral o son promovidos desde la cooperación internacional. En cualquier caso y considerando que los costos de los acuerdos son más altos para los consensos estratégicos como los descritos en el párrafo anterior, ante la ausencia de consensos básicos sobre políticas integrales de calidad, ante la falta de coincidencias acerca de la distribución de recursos o frente a fuertes discrepancias ideológicas, la dispersión de los costos de transacción en programas acotados –para los cuales no son necesarios acuerdos globales sino que alcanza con obtener acuerdos técnicos y que distribuyen recursos en menor escala– podría explicar la frecuencia con que estos programas son mencionados por los actores, como ejemplos del uso de los resultados de las evaluaciones.

El presente estudio no ha registrado la práctica sistemática y extendida de realizar evaluaciones de impacto entre los países que ha abarcado, aunque se registra el uso de resultados de pruebas a gran escala para evaluar un programa de Bachillerato Internacional y la extensión de la jornada escolar en educación básica en dos países que formaron parte de la investigación. Es probable que la baja frecuencia de uso para el monitoreo y la evaluación de programas se explique tanto por las características que asumen los datos de las evaluaciones –oportunidad, pertinencia, validez, cobertura– como por el hecho de que las evaluaciones de impacto no forman parte de las prácticas generalizadas de análisis de políticas en los espacios estatales de producción de información.

En relación con los usos curriculares, la diferencia está marcada por la existencia o no de una política nacional de estándares, entendiéndose por tal a aquella que establece parámetros de logro, junto con su revisión periódica, de acuerdo con los resultados obtenidos por las evaluaciones periódicas. El uso de resultados de evaluación y currículo depende del diseño de la política curricular. Otros usos, como la revisión parcial de áreas específicas del currículo basadas en pruebas internacionales como PISA o CIVIC, no se apoyan estrictamente en resultados, sino en los modelos de organización de competencias que presentan las matrices de estas pruebas.

Asimismo, se han registrado usos vinculados en forma más o menos directa con los modelos de consecuencias fuertes, como el de incentivos salariales a docentes, recursos dirigidos a establecimientos condicionados a resultados y la publicación de *rankings*. Estos empleos de los datos de evaluación tienen una característica particular en relación con la formulación de las políticas y es que constituyen resultados que están al servicio de una política ya formulada. Por lo tanto, no necesitan de una mediación o de la construcción de alternativas para su aplicación.

Las evaluaciones internacionales son puntos de referencia importantes para diversos usos. En primer lugar, sobresalen los usos comparados de las pruebas PISA, si bien estos usos acumulan numerosas críticas respecto de la validez de la comparación. Evidentemente, este aspecto conlleva riesgos en el uso, entre los que se cuentan restar legitimidad a las evaluaciones estandarizadas y simplificar el debate político. Sobresalen también los usos comparativos entre algunos estados subnacionales que, ya sea por razones políticas o estratégicas, tienen interés en buscar referencias fuera del espacio nacional, así como procuran diferenciarse de sus pares locales. Un segundo uso que emerge de los casos nacionales sobre PISA es el uso iluminativo o de producción de conocimiento (Weiss, 1989 y 1998), a partir del cruce de niveles de logro con factores asociados. Más allá de los debates, los productores y los usuarios especializados valoran la calidad técnica de las pruebas PISA. Otro uso que sobresale y se aplica a estas pruebas lo constituye el hecho de que funcionan como referencia para revisar la selección de conocimientos en las áreas de Lenguaje y Matemática.

Menos referencias se han recogido de las evaluaciones realizadas en el marco del LLECE, en parte, porque son pruebas cuya aplicación es más espaciada en el tiempo, en este sentido, la continuidad y comparabilidad a lo largo de los años es central en la valoración positiva de PISA. Por otra parte, los operativos SERCE y TERCE se han aplicado hace varios años y el ERCE, al momento de realizar el estudio, aún estaba en procesamiento.

Tal vez uno de los aportes principales de este estudio consista en la manera de articular los distintos factores y dimensiones ya conocidos, en una explicación que coloca el énfasis en el papel estructurador de las funciones generales del modelo evaluativo, en su diseño institucional y en su capacidad de estabilizar las tensiones generadas en las disputas por la regulación de los sistemas educativos.

No es el modelo de altas implicancias de por sí el que asegura el uso propiamente dicho de los resultados, sino la naturaleza, el tipo de implicancias y sobre quiénes se ejercen las mismas. En otras palabras, el éxito de la evaluación es posible cuando obliga a los responsables políticos a movilizar recursos presupuestarios, técnicos y de coordinación y cuanto proporciona las condiciones para que los actores escolares movilicen recursos pedagógicos (y, entre estas condiciones, se cuentan los recursos materiales y las capacidades técnicas en el marco de políticas pedagógicas).

La ligazón entre los resultados obtenidos y las políticas de calidad no necesariamente deben pensarse como acciones sobre los actores, en el sentido más común de premios y castigos, sino en términos de asignación y priorización de recursos, con una información transparente para todos acerca del cumplimiento de las responsabilidades de cada agente, combinados con un mayor conocimiento de “lo que funciona y lo que no funciona”. La experiencia acumulada en estas últimas décadas sobre resultados de evaluación y los estudios comparados ha dejado un repertorio de políticas de mejora que pueden ser consideradas como referencias. Sin embargo, para los propósitos de este estudio, pareciera ser especialmente relevante que el diseño de lo que se ha designado como ligazón entre evidencia y política esté acordado y diseñado desde un principio y que estos acuerdos se revisen periódicamente. Esto supone un consenso previo sobre el modelo de políticas de calidad, el enfoque sobre cumplimiento de derechos que las enmarca, las condiciones necesarias y progresivas para alcanzar el cumplimiento de esas políticas y la orientación de la asignación de recursos. Supone, en definitiva, aliviar a los dispositivos de evaluación de la carga desproporcionada de constituirse en arena de discusión de una agenda irresuelta.

8. Recomendaciones

A continuación, se presentan recomendaciones orientadas a fortalecer el uso de los resultados de las evaluaciones estandarizadas a gran escala. Se trata de orientaciones estratégicas, dado que las de carácter técnico u operativo y centradas en los casos pueden leerse en los Informes Nacionales.

Anclar los usos de los resultados en acuerdos sobre las funciones de los sistemas de evaluación

Como se ha señalado en el presente informe regional, en algunos países las evaluaciones y sus usos están sometidos a las tensiones que surgen acerca de definir cuáles deberían ser las funciones de la evaluación. Anclar los usos en acuerdos mínimos sobre el papel de la evidencia, que dejen transitoriamente por fuera las tensiones por desacuerdos, permitiría avanzar en instalar y consolidar los usos. Si estos acuerdos abarcaran un abanico de usos concretos que involucren a distintos actores –aun con distintas perspectivas acerca de la evaluación– ello sin dudas contribuiría a la construcción de la legitimidad de los resultados y de sus usos. El empleo de la información coloca en valor a los sistemas educativos que precisamente proveen esa información.

En relación a los usos previstos por las normativas y marcos orientadores, se observa que estos a menudo presentan una formulación con propósitos argumentales. Si bien estos propósitos no son irrelevantes en tanto que explicitan orientaciones y fijan una posición, el estudio muestra que, con diferentes énfasis, no es el nivel enunciativo de usos de resultados lo que diferencia los usos previstos de los usos efectivos, sino el entramado normativo y la definición estatutaria de las regulaciones, junto con la manera en que los diseños institucionales asimilan las tensiones sobre las funciones de la evaluación.

La formulación de prescripciones y el control de cumplimiento son de más simple resolución en el modelo que emplea funciones de altas implicancias, porque como se ha señalado a lo largo del documento, la decisión preexiste a la medición. Sin embargo, lo que pone de manifiesto el presente estudio es la importancia acerca de la existencia de un nexo entre los resultados y sus consecuencias, independientemente del tipo de implicancias que determine el modelo evaluativo.

Proyectar de manera progresiva e incremental los usos de las evaluaciones en el mediano plazo

Una recomendación que surge en relación con la dimensión estatutaria, para aquellos países que definan reforzar usos, por ejemplo, para atender desequilibrios de calidad manifiesta entre estados subnacionales, regiones, escuelas o poblaciones, sería la de precisar responsabilidades y priorizar recursos tanto presupuestarios como técnicos. Los modelos de evaluación que asumen funciones de bajas implicancias pueden asumir estos compromisos al vincular políticas de mejora con resultados esperados y con los ajustes correspondientes en caso de que los resultados no se alcancen. No se trata de desconocer los argumentos que se han recogido en varios Informes Nacionales de este estudio acerca de la dificultad de vincular políticas o programas específicos con resultados de evaluaciones, no obstante, si el objetivo es mejorar condiciones y reducir brechas de equidad, el empleo de distintas fuentes de información entre las que se encuentren los resultados de las evaluaciones a gran escala contribuye a la identificación de las barreras que perduran en el cumplimiento de dichos objetivos.

Los países que, como Argentina, se orientan por un modelo deliberativo en el que la información alimenta la construcción de acuerdos, podrían asumir un modelo incremental de logros y de responsabilidades asociado con planes plurianuales, acompañados de planes de evaluación que cuenten con varios años por delante. En este punto, puede resultar útil la figura de rendición de cuentas cooperativa

enunciada en este documento, en el sentido de concebir un tipo de rendición de cuentas que identifique responsabilidades compartidas y acuerdos mínimos acerca de su cumplimiento entre los actores involucrados. En este caso, los acuerdos y los compromisos deberían revisarse periódicamente de manera incremental.

Mantener continuidad en las mediciones

La continuidad de las mediciones, sus usos, las reformulaciones necesarias y una proyección incremental del uso de los resultados a partir de los acuerdos alcanzados, pareciera ser importante a juzgar por la experiencia de los países con sistemas estables y usos afianzados, como Brasil, Chile y Uruguay. En este sentido, puede ser útil el ejemplo de la evaluación PISA, que, si bien es objeto de múltiples críticas, también es valorada por su continuidad y calidad. La incorporación de planes plurianuales de evaluación y el uso de índices complejos que puedan revisarse periódicamente, como es el caso del IDEB en Brasil, a pesar de sus limitaciones y de sus efectos excesivamente simplificadores pareciera ser una vía operativa para promover el uso de la evidencia.

Alinear usos con las características técnicas de las pruebas

Además de anclar usos con acuerdos, se considera importante alinear usos con las características técnicas de las pruebas y con el nivel de gestión en el que se emplean. Los usos de las pruebas muestrales y censales no son equivalentes en todos los niveles de gestión. Por lo tanto, puede ser de utilidad anticipar los usos de los resultados cuando se discute un plan anual. Esta estrategia ha sido planteada por algunas organizaciones no gubernamentales usuarias de la información. Lo que se procura es concebir la comunicación de la información como un corpus con múltiples niveles y formatos de salida.

Definir políticas de uso de los resultados

El estudio pareciera aportar como conclusión que un plan plurianual que incluya niveles de logro académico entre sus metas expresa el uso efectivo de los resultados, cuando es parte de un diseño de políticas que establece una conexión entre los resultados y las políticas de calidad, es decir, cuando esta conexión antecede a las evidencias y son las evidencias las que desencadenan las políticas; también cuando se dispone de una logística institucional y de recursos que puedan ser movilizados bajo la responsabilidad de los organismos públicos sectoriales. Estas tres condiciones pueden ser aplicadas tanto en el caso de modelos con altas implicancias como en el caso de modelos de bajas implicancias, como lo son aquellos centrados en la evaluación formativa.

También es importante, como punto de partida, no presuponer la continuidad entre evidencia y acción sin asumir que los resultados no tienen posibilidad de usos si existe un vacío de políticas, a no ser que su destino sea la responsabilización unilateral de unos actores hacia otros. Al mismo tiempo, es pertinente reconocer que el empleo de los resultados tiene características y exigencias diferentes según los niveles de gestión y, en el caso específico del uso entre los niveles escolares y de aula, la utilización de resultados debe ser encuadrada en políticas pedagógicas.

Distinguir entre audiencias y usuarios

A partir de lo afirmado anteriormente, se deriva la relevancia de establecer una distinción entre audiencia y usuario, con el fin de definir una política de uso, en tanto las y los usuarios tienen responsabilidades específicas dirigidas desde las evidencias, así como también requerimientos de interpretación y de procesamientos diferentes. Definir usuarios supone reconocerles un papel específico en la política o en los programas, que debe ser informado claramente a partir de las evidencias proporcionadas por las evaluaciones. Sobre este tema, un

aporte del estudio en su conjunto es poner de manifiesto que existe una limitación al concebir la comunicación de los resultados en términos de audiencias, en tanto, se sugiere la conveniencia de desplazar la reflexión hacia la figura de las y los usuarios de los datos recabados a partir de la aplicación de las evaluaciones.

Fortalecer las capacidades técnicas de los estados subnacionales y de los usuarios locales

Por último, el estudio es convergente en señalar la importancia del fortalecimiento de las capacidades técnicas de los estados subnacionales, de los municipios y de las unidades de gestión territorial para contribuir a robustecer los usos de los datos de evaluación. Este fortalecimiento debería darse en el marco de las funciones asignadas a estos agentes específicos dentro de la política de evaluación.

Bibliografía

Backhoff, S. y Contreras Roldán. (2014). “Corrupción de la medida” e inflación de los resultados de enlace. *Revista Mexicana de Investigación Educativa (RMIE)*, 19, 63, pp. 1267-1283. México.

Barroso, L. y Carvalho, L.M. (2011). Apontamentos sobre os “novos modos de regulação” à luz de estudos sobre as relações entre conhecimento e política”. *Proposta Educativa*, 2 (36), pp. 9-36. FLACSO, Buenos Aires. Recuperado de <http://propuestaeducativa.flacso.org.ar/revista/indice-n36/>

Bolívar, A. (1996). El lugar del Centro Escolar en la política curricular. Más allá de la reestructuración y de la descentralización. En Pereyra, M. *et al.* (comp.). *Globalización y descentralización de los Sistemas Educativos. Fundamentos para un nuevo programa de educación comparada*. Barcelona: Pomares-Corredor.

Cabal, J. et al. (2018). Programa de Incentivos Docentes: impactos y recomendaciones en el corto plazo. Bogotá: Departamento Nacional de Planeación.

Cox, C. y Meckes, L. (2016). International large-scale assessment studies and educational policy-making in Chile: contexts and dimensions of influence. *Research Papers in Education*, 31:5, pp. 502-515.

Cueto, S. (ed.), (2005). *Uso e impacto de la información educativa en América Latina*. Santiago de Chile: PREAL.

Galas, C. et al. (2020). *Estado del arte sobre uso de resultados de evaluaciones en América Latina*. Buenos Aires: IIPE UNESCO.

IIPE UNESCO (2019). *Working Papers Use of learning assessment data in education policy-making (by Ieva Raudonyte)*. París: IIPE UNESCO.

INEED (s/f). *Políticas hacia el sector docente. Estudio analítico multipaís. Aportes para la elaboración de políticas educativas en el Uruguay* (por Flavia Terigi coord., Gabriela Diker,

Bárbara Briscioli y Carolina Lifschitz). Montevideo: Ministerio de Educación y Cultura.

Lindblom, C. (1994). *La investigación social para la elaboración de políticas: quién la necesita y para qué*. Gestión y Política, vol., III, núm. 2, pp. 253-291

Majone, G. (1997). Evidencia, argumentación y persuasión en la formulación de políticas. México: Fondo de Cultura Económica.

Maroy, C. (2008). Vers une régulation post-bureaucratique des systèmes d'enseignement en Europe? Sociologie et sociétés, 40, (1), 31-55. <https://doi.org/10.7202/019471ar>

Marradi, A. et al. (2007). Metodología en Ciencias Sociales. Buenos Aires. Emecé.

Morduchowicz, A. (2010). El federalismo fiscal-educativo argentino. En Andrade de Oliveira, D. et al. Políticas educativas y territorios modelos de articulación entre niveles de gobierno. Buenos Aires: IIEP UNESCO.

Morduchowicz, A. (2019). El financiamiento educativo argentino. Propuesta Educativa, Año 28, núm. 52, noviembre 2019, pp. 11-23.

Ravela, P. (2001). ¿Cómo presentan sus resultados los Sistemas Nacionales de Evaluación Educativa en América Latina? Santiago de Chile: PREAL.

Ravela, P. (2006). Para comprender las evaluaciones educativas. Fichas Didácticas. Santiago de Chile: PREAL.

Ravela, P. et al. (2007). Las evaluaciones educativas que América Latina necesita. Resultados de grupo de trabajo, PREAL-GRADE.

Silva, M. et al. (2012). El uso de los resultados de las evaluaciones del INEE por las autoridades educativas. México: INEE.

Terigi, F. (2004) La enseñanza como problema político. En Diker G. y Frigerio G. (2004) La transmisión en las sociedades, las instituciones y los sujetos: un concepto de la educación en acción. Centro de Publicaciones educativas y material didáctico. Buenos Aires.

UNESCO (2017). Rendir cuentas en el ámbito de la educación: cumplir nuestros compromisos, GEM Report. Publicado en 2017 por la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. Accesible en <https://en.unesco.org/gem-report/>

Venegas, M. *et al.* (Coords.), (2017). El profesorado frente a las políticas de rendición de cuentas: ¿qué redefinición(es) de su profesionalismo y de su responsabilidad social? *Revista de Sociología de la Educación (RASE)* 2017, vol. 10, Número monográfico. n° 3.

Weiss, C. (1979). The many meanings of research utilization. *Public Administration Review*, 39(5), 426-431. doi:10.2307/3109916

Weiss, C. (1998). Have we learned anything new about the use of evaluation? *American Journal of Evaluation*, 19(1), 21-33. Obtenido de <https://pdfs.semanticscholar.org/976f/28dfe-161d086b2605af6c3e364a7bf8bd991.pdf>

Anexos

Anexo 1: Criterios de selección de actores entrevistados

En una primera instancia se determinaron perfiles de dos universos: funcionarios y funcionarias gubernamentales, exfuncionarios o exfuncionarias, por un lado, e integrantes de organizaciones de la sociedad civil, por el otro. En las áreas gubernamentales, se definieron tres tipos genéricos: niveles de toma de decisión, áreas usuarias de información y áreas de evaluación (estas últimas, entendidas como productoras del tipo de información cuyo uso es objeto de este estudio). Luego, para cada uno de los subuniversos de las áreas gubernamentales se decidió entrevistar, a la vez, dos tipos de perfiles, uno de carácter político y otro de carácter técnico. El carácter “político” o “técnico” se identificó, según el país, por el modo de acceso al cargo o por la naturaleza de la función que desempeña.

En el caso de las áreas usuarias de información, se incluyeron áreas de planeamiento y áreas que se denominaron genéricamente “operativas”, entendiendo por tales a las unidades de gestión de la administración central sectorial cuyas funciones, de acuerdo con las especificaciones de los organigramas de sus respectivas administraciones, permitiera considerarlas como posibles usuarias de información de resultados de la evaluación. Inicialmente se pensó en formación y profesionalización docente, currículum y unidades a cargo de programas especiales. En esta misma categoría, se incluyó a las unidades que tuvieran responsabilidades directas sobre la supervisión, la fiscalización o el apoyo de establecimientos educacionales.

Para cada uno de los tres países federales, se decidió seleccionar dos estados subnacionales y, para el caso de Brasil, un municipio. Se seleccionaron entidades subnacionales que,

de acuerdo con información proporcionada por informantes clave, reportaran al menos uno de los siguientes dos rasgos: un uso intensivo de los resultados de las evaluaciones aplicadas nacionalmente, el desarrollo de pruebas, módulos o muestreos propios de evaluación de aprendizajes a gran escala (nacionales o internacionales).

Para Chile, Ecuador y Uruguay, se seleccionaron niveles de gestión centrales. En el caso de Chile, se entrevistó a responsables de los servicios locales de educación que administran y supervisan escuelas en territorios constituidos por un grupo de comunas.

En cuanto a las organizaciones de la sociedad civil, se procuró identificar organizaciones que, de acuerdo con la información disponible sobre su actuación pública, pudieran ser caracterizadas como usuarias de información proporcionada por las evaluaciones de aprendizajes a gran escala, que tuvieran incidencia en el debate de políticas educativas –o en los debates sobre el uso de evaluación–, que desempeñaran actividades de *advocacy* en educación, u oficinas de organismos internacionales que tuvieran un papel activo en cada escena educativa nacional. Asimismo, se incluyó a representantes gremiales por el papel que estos han desempeñado históricamente en el debate sobre políticas educativas y porque a menudo los resultados de las evaluaciones en la región han sido parte de políticas de regulación del desempeño de la docencia (Caballero *et al.*, 2018; Ravelo, *et al.*, 2007; Silvia, *et al.*, 2012).

Durante una segunda instancia, la definición general de perfiles descrita precedentemente fue adaptada para el estudio de caso de cada país. Los criterios de adaptación fueron cuatro. En primer lugar, las variaciones del diseño organizacional del sector educación, por ejemplo, la dependencia de la educación básica y de la educación media superior respecto de unidades de gestión diferentes, o alguna otra variación que indicara incorporar algún perfil específico que no estuviera incluido en la definición general. En segundo lugar, la

existencia de actores no gubernamentales especializados en educación, cuya actuación en el escenario nacional sugiriera interés para incorporarlos y que no estuvieran comprendidos en las categorías precedentes. Un tercer criterio de adaptación se puso en juego cuando en el diseño organizacional nacional se detectó la existencia de usuarios de información de la evaluación a gran escala externos a los organismos sectoriales, como el poder legislativo u organismos de contralor. Finalmente, también se aceptó un margen de variabilidad en el número de exfuncionarios y exfuncionarias a entrevistar, al considerar los procesos recientes de recambio en las administraciones gubernamentales o cuando se detectaran perfiles que por su papel y trayectoria en el desarrollo del dispositivo de evaluación de aprendizajes de gran escala se ameritara incorporar su perspectiva.

Anexo 2: Características de las evaluaciones por país

(Nacionales e internacionales, con excepción de los casos donde se indica)

Argentina

Fuente: Informe Nacional.

Descriptores	Tipo de evaluación de aprendizajes		
	Aprender (Nacional)	ERCE (Regional)	PISA (Internacional)
Objetivos	Obtener y generar información oportuna y de calidad para conocer mejor los logros alcanzados y los desafíos pendientes en torno a los aprendizajes de los estudiantes para contribuir a procesos de mejora educativa continua.	Mide logros de aprendizaje en relación con los objetivos de aprendizaje que son comunes en la región según los planes de estudio de cada país.	Describir la situación de la educación escolar en los países y promover su mejoramiento. Busca evaluar la capacidad de los estudiantes de entender y resolver problemas de la vida cotidiana y social.
Población a la que se aplica	Estudiantes de 3° grado/año y del último año de cada nivel obligatorio.	Estudiantes de nivel primario de América Latina y el Caribe de 3° y 6° grado/año.	Estudiantes de educación formal de 15 años de edad, independientemente del grado o año que cursen
Año de aplicación	2016, 2017, 2018 y 2019	2019	2018
Alcance	Muestral, en 2016, en el 3er grado/año. Luego fue censal.	Muestral	Muestral
Periodicidad	Anual (excepto para primaria, en 2019, y para secundaria, en 2018).	Variable	Triannual

Dominios o áreas de conocimiento evaluados	Lengua, Matemática, Ciencias Naturales y Ciencias Sociales. En 2019, se incluyó Educación ciudadana.	Lenguaje (lectura y escritura) y Matemáticas. La prueba del área de Ciencias Naturales es rendida solo por los estudiantes de sexto grado de educación primaria.	Matemática, Lectura y Ciencias Naturales. En 2018, el foco estuvo puesto en Lectura.
Tipo de escuelas	Primarias y secundarias de sector estatal y privado.	Primarias del sector estatal y privado.	Secundarias del sector estatal y privado.
Organismo responsable	Secretaría de Evaluación del Ministerio de Educación de la Nación.	Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE), UNESCO.	Organización para la Cooperación y el Desarrollo Económicos (OCDE).
Información de contexto en cuestionarios complementarios	El cuestionario complementario recoge información sobre clima escolar, autopercepción del estudiante, prácticas educativas, usos de tecnologías, contenidos sobre Educación sexual integral trabajados y acciones de capacitación realizadas por docentes. Son completados por directivos, docentes y estudiantes.	Incluye cuestionarios que completan diferentes miembros de la comunidad educativa: estudiantes, familias, docentes y directores, con el fin de conocer los contextos en los cuales los aprendizajes de los estudiantes tienen lugar.	Los cuestionarios complementarios de las pruebas analizan aspectos vinculados con la oferta educativa, las características de los sistemas, la organización interna de las instituciones escolares, las prácticas de enseñanza, las percepciones de estudiantes, docentes y directivos, y el contexto socio-económico y cultural.

Brasil

(Evaluaciones únicamente nacionales)

Fuente: Informe Nacional.

Pruebas bianuales que incluyen un cuestionario de antecedentes			
Público objetivo	Abrangencia	Formulación	Áreas de conocimiento / Materias evaluadas
Guardería y preescolar de la educación inicial	Escuelas públicas (muestral) - Estudio piloto	BNCC	--- (recopilación de cuestionarios con secretarios, directivos y docentes)
2° año de la Enseñanza Fundamental	Escuelas públicas (muestral) Escuelas privadas (muestral)	BNCC	Lengua portuguesa y matemáticas
5° y 9° año de la Enseñanza Fundamental	Escuelas públicas (censitario) Escuelas privadas (muestral)	Matriz de Referencia	Lengua portuguesa y matemáticas
9° año de la Enseñanza Fundamental	Escuelas públicas (muestral) Escuelas privadas (muestral)	BNCC	Ciencias naturales y ciencias humanas
	Escuelas públicas (censitario) Escuelas privadas (muestral)	Matriz de Referencia	Lengua portuguesa y matemáticas
Indicador	Abrangência	Composição	Periodicidade

<p>IDEB –Índice de Desenvolvimento da Educação Básica</p>	<p>El indicador se calcula para la educación primaria (años iniciales de la enseñanza fundamental), la educación secundaria baja (años finales de la enseñanza fundamental) y la educación secundaria alta (enseñanza media). El IDEB se calcula para el país, los estados y los municipios; las redes pública y privada; y para las escuelas públicas.</p>	<p>Desempeño (resultado de las pruebas estandarizadas de lengua portuguesa y matemáticas) y rendimiento escolar (aprobación). Un mayor rendimiento y una mayor aprobación se reflejan en un mayor IDEB.</p>	<p>Bianual</p>
<p>Metas del IDEB</p>	<p>Ídem</p>	<p>Calculado utilizando la competencia media de los países de la OCDE en 2005 como referencia para 2021</p>	<p>Bianual. Calculadas hasta 2021.</p>

Chile

(Únicamente evaluaciones nacionales)

Fuente: Informe Nacional.

Descriptores	Nombre de la evaluación		
	SIMCE	Pruebas nacionales	Evaluación progresiva
Objetivos	<ul style="list-style-type: none"> - Contribuir al mejoramiento de la calidad y equidad de la educación. - Evaluar el logro de los Objetivos fundamentales y los Contenidos mínimos obligatorios del Marco curricular. - Monitorear el nivel de desempeño de niños y jóvenes mediante la evaluación permanente. - Focalizar los recursos y el apoyo. - Examinar el desempeño de las escuelas para identificar fortalezas y debilidades. - Entregar información de los resultados a la comunidad educativa. - Garantizar la rendición de cuentas de escuelas o educadores. - Planificar reformas de política educativa. 	<ul style="list-style-type: none"> - Monitorear los logros de aprendizaje y evaluar la efectividad de las políticas del Estado. - Reforzar los compromisos del Estado y de los actores educativos con las mejoras en la calidad. - Focalizar los recursos y el apoyo. - Orientar el trabajo de otras instituciones, como las formadoras de docentes y los organismos encargados. 	<ul style="list-style-type: none"> - Aportar información oportuna y específica a los establecimientos, respecto de los avances en el desarrollo de los aprendizajes del alumnado. - Entregar orientaciones pedagógicas a los establecimientos. - Implementar acciones pedagógicas que impacten en la mejora de los aprendizajes de sus estudiantes. - Levantar información respecto al diagnóstico, monitoreo y trayectoria del estudiantado en los establecimientos.

Población a la que se aplica	CINE 1; 4° Grado CINE 1; 6° Grado CINE 1; 8° Grado CINE 2; 2° Grado	CINE 1; 2° Grado CINE 1; 8° Grado CINE 2; 3° Grado	CINE 1; 2° Grado CINE 2; 1° Grado
Alcance	Censal	Muestral	Muestral
Periodicidad	Anual	Anual	Anual
Dominios evaluados	-Lectura -Matemáticas -Ciencias medioambientales -Historia, Geografía y Ciencias Sociales -Escritura	-Lectura -Educación ciudadana -Educación física y salud -Inglés -Educación Media Técnico Profesional (CINE 2; 4° Grado)	-Lectura -Matemáticas
Tipos de escuelas	Públicas y privadas.	Públicas y privadas.	Públicas y privadas.
Organismo responsable	Agencia de Calidad de la Educación.	Agencia de Calidad de la Educación.	Agencia de Calidad de la Educación.
Información de contexto	Cuestionarios de: Estudiantes Docentes Apoderados(as)/hogar Escuela		

Ecuador

Fuente: Informe Nacional.

Descriptores	Nombre de la evaluación			
	Ser Estudiante	Ser Bachiller	SERCE - TERCE - ERCE	PISA-D
Objetivos	<p>Evaluar las destrezas y saberes de los estudiantes de 4°, 7° y 10° de Educación general básica (EGB) en los campos de: Matemática, Lengua y Literatura, Ciencias Naturales y Estudios Sociales.</p> <p>Medir los saberes a través de los estándares de aprendizaje establecidos por el MINEDUC.</p> <p>Promover una educación de excelencia a través de la evaluación integral al Sistema l de Educación y sus componentes.</p>	<p>Explorar las destrezas con criterio de desempeño que los sustentantes deben alcanzar al culminar la educación obligatoria que son necesarias para el desenvolvimiento exitoso como ciudadanos y ciudadanas, y para afrontar estudios de educación superior; en concordancia con lo que establece el perfil de salida del bachiller ecuatoriano.</p> <p>Formar parte de la nota de grado para concluir el Bachillerato y del proceso de admisión a la Educación Superior.</p>	<p>Generar información relevante sobre el logro de aprendizaje y otros indicadores educativos, que dan cuenta de diversos aspectos relacionados con la calidad de la educación en un sentido integral y responde a los desafíos de la Agenda de Educación 2030.</p>	<p>Lograr que PISA sea más accesible y relevante para países de ingreso medio y bajo.</p> <p>Evaluar hasta qué punto las y los estudiantes que están terminando la educación obligatoria de Ecuador son capaces de aplicar sus conocimientos en situaciones de la vida real y están preparados para participar plenamente en la sociedad.</p>

Población a la que se aplica (años)	4°, 7°, 10° de EGB.	3° de BGU.	4° y 7° de EGB.	Estudiantes regulares de 15 años.
Alcance (censal -muestral)	Muestral	Censal	Muestral	Muestral
Periodicidad	Anual	Anual	Cada 6 años	Cada 3 años
Dominios evaluados	Matemática, Lengua y Literatura, Ciencias Naturales, Ciencias Sociales.	Matemática, Lengua y Literatura, Ciencias Naturales, Ciencias Sociales.	Matemática, Lectura y Escritura y Ciencias de la Naturaleza.	Lectura, Matemáticas y Ciencias.
Tipos de escuelas (públicas -privadas)	Fiscales, fiscomisionales, privadas y municipales.	Fiscales, fiscomisionales, privadas y municipales.	Fiscales, fiscomisionales, privadas y municipales.	Fiscales, fiscomisionales, privadas y municipales.
Organismo responsable	INEVAL	INEVAL	LLECE	OCDE
Información de contexto	Sí	Sí	Sí	Sí
Relación con el currículo nacional	Sí	Sí	Sí	No

México

(Evaluaciones únicamente nacionales)

Fuente: Informe Nacional.

Descriptor	Estándares nacionales	EXCALE	ENLACE	PLANEA
Año del primer levantamiento	1998	2005	2006	2015
Población a evaluar	Estudiantes de los seis grados de primaria y de los tres de secundaria.	Estudiantes de 3° de preescolar y primaria, 3° de educación secundaria y último año de media superior.	Estudiantes de 3° a 6° de primaria. Se extendió a los tres grados de la enseñanza secundaria y, más tarde, al 3° grado de la educación media superior.	Estudiantes que terminan el 3° grado de preescolar, el 6° de primaria, el 3° de secundaria y el último grado de educación media superior.
Periodicidad	Se tiene registro de que estas pruebas se aplicaron en los ciclos escolares 1997-1998, en primaria, y 1999-2000, en secundaria.	Cada cuatro años.	Cada año.	ELSEN: cada tres años. ELCE: cada año. EDC: cada año.
Institución responsable	SEP	INEE	SEP	INEE, SEP, Autoridades estatales

Alcance	Sus resultados se presentaban con altos niveles de agregación, es decir, con referencia a grupos muy amplios de estudiantes.	Muestral	Censal	Muestral por alumnos (ELSEN). Censal por escuelas (ELCE). Diagnóstica censal.
Características generales	Su objeto era conocer el nivel de desempeño de las y los estudiantes mexicanos. No tenían una cobertura suficiente del currículo. Adolecían de problemas técnicos.	Su objetivo era ofrecer a las autoridades educativas federales y locales, así como a la opinión pública, información válida y confiable del desempeño del sistema educativo nacional en su conjunto. Prueba alineada con el currículo nacional.	Su propósito era obtener un diagnóstico del aprendizaje adquirido por cada estudiante. No tenía niveles altos de control durante su relevamiento. Prueba alineada con el currículo nacional.	Su objetivo es conocer la medida en que el estudiantado logra el dominio de un conjunto de aprendizajes esenciales en diferentes momentos de la educación obligatoria. Se centra en la rendición de cuentas a nivel nacional. ^{47*}
Año del último relevamiento	2003	2009	2014	Suspendida hasta nuevo aviso por la pandemia del virus COVID-19.

Uruguay

Fuente: Informe Nacional.

Descriptoros	Nacionales			Internacionales	
	Formativa	Adaptativa	Aristas	PISA	LLECE
Objetivos	Promover instancias de reflexión entre los actores educativos y contribuir a la mejora de la calidad de la educación.	Promover instancias de reflexión entre los actores educativos y contribuir a la mejora de la calidad de la educación. Valorar la dispersión de aprendizaje intergrado.	Monitorear los logros del sistema educativo en sus distintas dimensiones; brindar información útil para evaluar la garantía del derecho a la educación del estudiantado.	Evaluar el nivel de conocimientos adquiridos y habilidades desarrolladas que permita a las y los estudiantes participar plenamente en la sociedad como ciudadanos y ciudadanas críticos y comprometidos.	Conocer e informar sobre los niveles de desempeño de las y los estudiantes en las áreas de conocimiento seleccionadas.
Población a la que se aplica (años)	3° y 6° de Educación Primaria y 3° de Educación Media.	3° y 6° de Educación Primaria.	3° y 6° de Educación Primaria y 3° de Educación Media.	Estudiantes de 15 años escolarizados en cursos posprimarios.	3° y 6° de Educación Primaria.
Alcance	Censal	Censal	Muestral	Muestral	Muestral
Periodicidad	Anual	Anual	Triannual	Triannual	Discontinuidad: 2006-2013-2019

Dominios evaluados	Lengua, Matemática y Ciencias Naturales.	Lengua, Matemática y Ciencias Naturales.	Lectura y Matemática.	Lectura, Matemática y Ciencias Naturales.	Lectura, Escritura, Matemática y Ciencias Naturales (solo en 6°).
Tipos de escuelas	Aplicación voluntaria. Públicas y privadas, urbanas y rurales.	Aplicación voluntaria. Públicas y privadas, urbanas y rurales.	Públicas y privadas, urbanas y rurales.	Públicas y privadas, urbanas y rurales.	Públicas y privadas, urbanas y rurales.
Organismo responsable	ANEP / DSPE / DIEE	ANEP / DSPE / DIEE	INEEd	DIEE	DIEE
Información de contexto (cuestionarios asociados)	No se incluye.	No se incluye.	Cuestionario del alumno sobre contexto y convivencia, habilidades socioemocionales (en 6° de Primaria y 3° de Media). Cuestionario de familia sobre contexto y convivencia (solo en Primaria). Cuestionarios a docentes y directores sobre contexto y convivencia.	Cuestionario de factores asociados.	Información sobre las dimensiones de la escuela, el aula y el contexto que contribuyen a los aprendizajes.



unesco

Instituto Internacional
de Planeamiento de
la Educación