



Organización
de las Naciones Unidas
para la Educación,
la Ciencia y la Cultura



Instituto Internacional de
Planeamiento de la Educación
• IIPE-UNESCO Buenos Aires
• Oficina para América Latina



ESTADO DEL ARTE
DE LA INVESTIGACIÓN
A LA POLÍTICA

Uso de Resultados de las Pruebas de Aprendizaje en el Diseño de las Políticas Educativas en América Latina

Cecilia Galas Taboada,
María Teresa Gutiérrez Cortés
y Eva María Hamilton Vélez

Marzo 2020

Advertencia

El uso de un lenguaje no discriminator 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 de la UNESCO. Sin embargo, 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 sea 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 desdoblamiento léxico 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.

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 2020 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 2020

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).



La serie de documentos *Estado del Arte de investigaciones de políticas educativas* se inscribe en el marco de las actividades del área de Investigación y Desarrollo de la Oficina para América Latina del Instituto Internacional de Planeamiento de la Educación (IIPÉ), cuya misión consiste en fortalecer las capacidades de los Estados Miembros de la UNESCO en la región para planificar y gestionar sus sistemas educativos.

La producción de estos documentos se articula específicamente con el proyecto Sistema de Información de Tendencias Educativas en América Latina (SITEAL). El SITEAL es un observatorio en línea que ofrece una base de políticas, normativas, investigaciones y estadísticas sobre la situación educativa en la región latinoamericana, a partir de las cuales se producen documentos que sistematizan e interpretan dicha información para su seguimiento.

El objetivo de los Estados del Arte es facilitar a los equipos técnicos de gobierno, investigadores, docentes y capacitadores, documentos que sistematicen y analicen investigaciones sobre ciertos temas relevantes para la planificación y gestión de la política educativa. A través de este tipo de estudio, se intenta generar conocimientos útiles, con la finalidad de identificar y resumir el nivel de desarrollo de cada tema, reconociendo dimensiones, categorías y focos, así como también los puntos de disidencia, las constantes, las tendencias y los desafíos actuales que emergen en cada caso. El propósito de estos estudios es aportar a la comprensión del tema seleccionado y, al mismo tiempo, enriquecer el debate educativo.

Lo que se espera es que, mediante sus distintos ejes, esta serie de documentos funcione como un conjunto de estudios que faciliten la traducción del conocimiento producido por el mundo académico hacia el campo de la política pública.

Contenido

1. Introducción.....	5
1.1 Objetivos y preguntas de orientación para la revisión de la literatura.	5
1.2 Alcance y metodología de la revisión	6
1.3 Estructura de la revisión.....	7
2. Breve recuento del desarrollo de los sistemas de evaluación en la región	7
2.1 Creación de los sistemas de evaluación: contexto y motivaciones	8
2.2 Primera oleada: aparición de evaluaciones de gran escala y creación de sistemas de evaluación.....	15
2.3 Segunda oleada: evolución institucional, consideración de factores asociados y devolución de resultados por niveles de logro	16
2.4 Tercera oleada: desarrollo de mediciones más complejas, incorpora- ción de elementos no cognitivos en la evaluación y preocupación por el uso de los resultados.....	17
2.5 Marcos institucionales de las unidades evaluadoras	18
3. Marco conceptual sobre usos de resultados de evaluaciones.....	22
3.1 Tipología de usos de evidencia.....	22
3.2 Propósitos de la evaluación.....	23
3.3 Interacción entre visiones y funciones de la evaluación y el uso de resultados	24
3.4 Teorías de cambio de las evaluaciones en gran escala	25
4. Uso de resultados de evaluaciones en políticas educativas e interven- ciones en la región.....	27
4.1 Definición de la agenda política	28
4.2 Planeación estratégica	29
4.3 Focalización de programas e intervenciones	29
4.4 Orientación del currículo y materiales	30

4.5 Incentivos docentes.....	30
4.6 Clasificación o ranking de escuelas.....	30
5. Factores que influyen en el uso (o no uso) de resultados de evaluaciones	31
5.1 Factores políticos.....	32
5.2 Elementos institucionales	34
5.3 Aspectos técnicos.....	36
5.4 Comunicación y difusión	39
6. Posturas frente a las evaluaciones estandarizadas.....	41
6.1 Teórico-ideológicas.....	41
6.2 Técnico-instrumentales	43
7. Conclusiones.....	48
8. Bibliografía.....	50

1. Introducción

Durante la década de los noventa se realizaron importantes esfuerzos para la construcción de sistemas de evaluación de gran escala en América Latina. A casi treinta años del inicio de su implementación, resulta importante examinar cuáles han sido los avances, aprendizajes y retos para definir la agenda política y académica de la siguiente década. A lo largo de los últimos veinte años, se han realizado estudios en profundidad al respecto de esta trayectoria de desarrollo y consolidación de los sistemas de evaluación en la región. Sobresalen los esfuerzos del Programa para la Promoción de la Reforma Educativa en América Latina y el Caribe (PREAL) durante la década del 2000, y más recientemente del Banco Mundial y el Centro de Implementación de Políticas Públicas para la Equidad y el Crecimiento (CIPPEC). La presente revisión de literatura busca contribuir a integrar -y donde resulte pertinente, a actualizar- los distintos hitos de la conversación en torno al uso de resultados de evaluaciones en la región para identificar los temas prioritarios en este momento.

1.1. Objetivos y preguntas de orientación para la revisión de la literatura

El propósito de esta revisión de literatura es explorar el estado del arte sobre el uso de resultados de evaluaciones de aprendizaje como instrumentos de mejora en los países de América Latina. El enfoque de este documento consiste en identificar patrones en el uso, elementos que influyen en los usos observados, así como las distintas posturas que actualmente existen en torno a las evaluaciones de gran escala.

Las síntesis y conclusiones que se presenten servirán para identificar temas específicos en los que será pertinente ahondar con el fin de reconocer factores y dinámicas que favorecen o inhiben el aprovechamiento de estos resultados para la mejora, considerando la complejidad del proceso educativo y de la elaboración de políticas, la propia historia de la región, así como los distintos actores involucrados.

Concretamente, las preguntas que se pretenden responder son las siguientes:

- ¿Cómo ha sido el desarrollo de los sistemas de evaluación en América Latina y cuáles han sido las principales motivaciones que han guiado su evolución?
- ¿Qué modelos de usos de resultados son referentes y han sido utilizados por países de la región?
- ¿De qué manera han influido los resultados de evaluaciones en las políticas e intervenciones en los países de la región?
- ¿Qué factores influyen en el uso (o no uso) que se hace de los resultados?
- ¿Qué posturas existen en la región, actualmente, frente a los sistemas de evaluaciones estandarizadas?

1.2 Alcance y metodología de la revisión

La selección de documentos para el análisis se hizo conforme a los siguientes criterios:

- Trabajos publicados en los últimos quince años.
- El contenido aborda explícitamente el tema del uso de resultados de evaluaciones. En este sentido, se descartaron descripciones detalladas de sistemas de evaluación, discusiones técnicas sobre las características de las pruebas y, por lo general, informes descriptivos de resultados. Excepciones a este último caso se hicieron cuando alguna sección del reporte planteaba sugerencias de uso de la información presentada.
- Los casos se refieren, principalmente, a experiencias en países de la Latinoamérica. Aunque se incluyen algunos documentos del Programa para la Evaluación Internacional de Alumnos (PISA) y de referentes internacionales en el uso de evaluaciones, se descartaron trabajos que abordaban el tema del uso a nivel internacional de manera general para favorecer una visión más enfocada en los aciertos y retos particulares de la región.
- Se incluyeron trabajos de investigación tanto teórica como empírica.
- De acuerdo con las preguntas de orientación, se incluyeron en la revisión diversos tipos de texto: publicaciones académicas, artículos de opinión, reportes institucionales, capítulos de libros y presentaciones en seminarios o conferencias.

Es de particular interés representar en este trabajo las distintas voces que participan en la conversación regional en torno a la pertinencia y el uso de evaluaciones de gran escala. Por esta razón, en algunos casos excepcionales, se incluyeron artículos que, sin cumplir con estándares convencionales de rigor técnico, presentan posturas relevantes para esta conversación.

La extensión de la literatura revisada se fue delimitando conforme se observaba redundancia en las referencias y se alcanzaba saturación en la argumentación de un tema. Aun así, como es natural en trabajos de esta índole, pueden haberse excluido obras relevantes que será necesario recuperar en estudios posteriores.

1.3 Estructura de la revisión

La revisión se estructura de manera temática, integrando los aportes y perspectivas de distintos autores acerca de un tema específico, de acuerdo con su posicionamiento al respecto.

En el primer apartado se presenta una síntesis de la evolución de los sistemas de evaluación en la región para enmarcar los hallazgos presentados en el resto del documento respecto a los usos de resultados, las consecuencias que han tenido sobre los sistemas educativos y la propia evaluación, así como las críticas y recomendaciones que se hacen desde distintos frentes.

El segundo apartado presenta una serie de referentes conceptuales que han informado la discusión sobre el uso de resultados en la región. Los distintos marcos abordan tanto el uso de evidencias e información para el diseño de políticas como las distintas lógicas de cambio que subyacen en los sistemas de evaluación.

El tercer apartado se dedica a describir usos de resultados de evaluaciones de gran escala observados en la región, presentando ejemplos y casos de países con distintas trayectorias en evaluación.

En el cuarto apartado se abordan algunos factores que influyen en la forma de utilizar los resultados, de nuevo, a partir de prácticas documentadas en los países latinoamericanos.

El quinto apartado recoge posturas de varios autores respecto a la pertinencia, implementación, uso y consecuencias de las evaluaciones de gran

escala, identificando puntos de controversia y consenso.

Finalmente, las conclusiones recogen las ideas principales que emergen de la revisión de la literatura y proponen líneas de investigación futura.

2. Breve recuento del desarrollo de los sistemas de evaluación en la región

Al finalizar las dos primeras décadas del siglo XXI, aproximadamente 65 millones de adolescentes y jóvenes asisten a la educación secundaria en América Latina. En 13 de los 19 países de la región este nivel de enseñanza ha sido considerado obligatorio de modo completo y en el resto, con excepción de Nicaragua, la obligatoriedad alcanza a sus primeros años (secundaria baja o ciclo básico, según la denominación utilizada en cada país). No obstante el esfuerzo de inclusión, todavía un 15% de la población que tiene entre 11 y 17 años no se encuentra asistiendo a un establecimiento educativo y una parte importante de la población adulta no alcanzó a completar este nivel de enseñanza (SITEAL, 2019).

Para comprender adecuadamente las características, enfoques y usos que se han dado, y se dan actualmente, a la evaluación y sus resultados, es necesario remontarse al contexto histórico en el que surgen los sistemas de evaluación en Latinoamérica y los principales momentos de su evolución. Reconocer la influencia del contexto político, educativo y técnico en el que se diseñaron las primeras evaluaciones ayuda a explicar algunos de los aciertos, retos, apoyos y controversias que se han generado en torno al uso de resultados de las pruebas estandarizadas.

En el documento “Las evaluaciones educativas que América Latina necesita” (Ravela, et al., 2007), los autores sintetizan de manera precisa el recorrido de los sistemas de evaluación educativa en la región desde la década de los noventa hasta el año 2007, distinguiendo el comportamiento entre países: unos que conservan sus sistemas con algunos cambios, otros con discontinuidad en las aplicaciones, y otros tantos reiniciando y replanteando los enfoques, muchas veces por motivaciones de tipo político.

En su libro *América Latina después de PISA. Lecciones aprendidas de la educación en siete países* (2015), Rivas enmarca el desarrollo de los sistemas de evaluación

dentro de un análisis de las tendencias de la política educativa en la región.

De la misma forma, Felipe Martínez Rizo y Guillermo Ferrer hacen un recuento de la aparición de evaluaciones estandarizadas nacionales en los países latinoamericanos, así como de su evolución a lo largo del tiempo (Martínez Rizo F., 2008; Martínez Rizo F. , 2016; Ferrer, 2006; Ferrer & Fiszbein, 2015).

A continuación se describen, de manera concisa, los principales hitos en el desarrollo de sistemas de evaluación en la región, y se presentan algunas interpretaciones del marco histórico-político en el que surgen.

2.1 Creación de los sistemas de evaluación: contexto y motivaciones

Si bien Brasil, Costa Rica y Chile habían incursionado en el tema de las evaluaciones a gran escala desde mediados de los años ochenta, las evaluaciones estandarizadas se implementaron en países de América Latina con mayor fuerza a lo largo de los años noventa (Martínez Rizo, 2008).

Esto supuso la creación o fortalecimiento de cuadros técnicos especializados en evaluación y la aparición o ampliación de funciones de instancias evaluadoras. En varios casos, la creación del sistema de evaluación como tal fue posterior al inicio de la aplicación de pruebas de gran escala.

Tabla 1. Inicio de la aplicación de evaluaciones estandarizadas en la región en el marco de la evaluación de la calidad educativa

Año de la primera evaluación de gran escala en el marco de un sistema de evaluación de la calidad educativa ¹	Países
1982	Chile ²
1986	Costa Rica
1990	Brasil
1991	Colombia, República Dominicana
1992	Guatemala
1993	Argentina, El Salvador
1996	Cuba, Ecuador, Nicaragua, Paraguay, Perú, Uruguay
1997	Bolivia, Honduras
1998	México, Venezuela ³
2000	Panamá

Elaboración propia a partir de "La evaluación de aprendizajes en América Latina" (Martínez Rizo, 2016).

¹ Desde la década de los sesenta, varios países habían aplicado pruebas de gran escala o estandarizadas para distintos fines. Sin embargo, no se enmarcaban propiamente dentro de sistemas de evaluación con enfoque en medir la calidad educativa.

² En 1968 se llevó a cabo la primera evaluación estandarizada en el país y en 1982 se aplicaron las primeras pruebas de lo que seis años después sería el Sistema de Medición de la Calidad de la Educación (SIMCE).

³ El levantamiento de 1998 ha sido el único ejercicio de evaluación a escala realizado en el país, hasta la fecha.

Un esfuerzo importante en la agenda evaluativa regional fue la creación del Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE) en 1994 con el triple objetivo de promover políticas basadas en evidencia a través de la producción de información sobre la calidad de la educación y factores asociados, desarrollar capacidades de evaluación educativa y servir como foro para el intercambio de ideas y mejores prácticas en educación (UNESCO, 2014). En 1997, los países miembros del LLECE aplicaron el Primer Estudio Regional Comparativo y Explicativo (PERCE) sobre logros de aprendizaje de lectura y matemáticas de alumnos de tercero y cuarto de primaria. Al PERCE siguieron tres estudios más en los cuales se alternaron los grados de aplicación - se cambió el cuarto grado por sexto- y se ampliaron los campos de conocimiento evaluados -se agregó la prueba de ciencias para alumnos de sexto de primaria- (Jiménez Moreno, 2016; UNESCO, 2014).

Tabla 2. Participación de países en las evaluaciones del LLECE

Estudio	Año de aplicación	Número de países participantes	Listado de países
PERCE	1997	13	Argentina, Bolivia, Brasil, Chile, Colombia, Costa Rica, Cuba, Honduras, México, Paraguay, Perú, República Dominicana y Venezuela
SERCE	2006	16	Argentina, Brasil, Chile, Colombia, Costa Rica, Cuba, Ecuador, El Salvador, Guatemala, México, Nicaragua, Panamá, Paraguay, Perú, República Dominicana y Uruguay ⁴
TERCE	2013	15	Argentina, Brasil, Chile, Colombia, Costa Rica, Ecuador, Guatemala, Honduras, México, Nicaragua, Panamá, Paraguay, Perú, República Dominicana y Uruguay ⁵ y el estado mexicano de Nuevo León.
ERCE 2019	2019	18	Argentina, Brasil, Bolivia, Chile, Colombia, Costa Rica, Cuba, Ecuador, Guatemala, Honduras, El Salvador, México, Nicaragua, Panamá, Paraguay, Perú, República Dominicana y Uruguay ⁶

Elaboración propia a partir de datos del LLECE.

Con la creación de PISA, de la Organización para la Cooperación y el Desarrollo Económicos (OCDE), en el año 2000, se cimentaron las políticas de evaluación estandarizada en la región, y esta prueba internacional se colocó en el centro de la conversación de la política educativa y de otros sectores, por ser una herramienta que permite establecer comparaciones a nivel regional y global.

⁴ Bolivia, Honduras y Venezuela no participaron en esta segunda aplicación, mientras que se sumaron Ecuador, El Salvador, Guatemala, Nicaragua, Panamá y Uruguay.

⁵ A pesar de haber participado en los dos estudios anteriores, Cuba decidió no realizar el tercer estudio. El Salvador tampoco participó, y se incorporó por primera vez Honduras.

⁶ Cuba, El Salvador y Bolivia se reincorporan al estudio. Es la primera evaluación internacional en la que participa Bolivia desde 1997.

Tabla 3. Incremento en la participación de países de la región en PISA

Año de aplicación	Número de países participantes	Número de países participantes de la región	Listado de países de la región
2000	32	5	Argentina, Brasil, Chile, México y Perú ⁷
2003	38	3	Brasil, México y Uruguay
2006	51	6	Argentina, Brasil, Chile, Colombia, México, Uruguay
2009	66	8	Argentina, Brasil, Chile, Colombia, México, Panamá, Perú y Uruguay
2012	57	8	Argentina, Brasil, Chile, Colombia, Costa Rica, México, Perú, Uruguay
2015	72	9	Argentina, Brasil, Chile, Colombia, Costa Rica, República Dominicana, México, Perú y Uruguay ⁸
2018	79	9	Argentina, Brasil, Chile, Costa Rica, República Dominicana, México, Panamá, Perú y Uruguay
PISA-D (2018)	8	5	Ecuador, Guatemala, Honduras, Panamá y Paraguay

Elaboración propia a partir de reportes de PISA

En su artículo “Is it worth it? Rationales for (non) participation in international large- scale assessments” (2019), Camilla Addey y Sam Sellar retoman de Steiner-Khamsi la agrupación de motivaciones por participar en evaluaciones internacionales en cuatro razones o dimensiones, y las desarrollan de la siguiente manera:

- Dimensión política: se enfoca en factores políticos a nivel nacional, regional e internacional para movilizar ideas y legitimar decisiones y políticas. Las evaluaciones son instrumentos para crear coyunturas y ventanas de oportunidad política. Se valora más la narrativa construida que las inferencias válidas a partir de los datos. También se relaciona con discursos

⁷ Chile y Perú implementaron la evaluación en 2002 y Argentina en 2003, (OCDE, 2002).

⁸ Venezuela sigue siendo país miembro del LLECE, sin embargo, no participó en la aplicación del ERCE 2019.

de rendición de cuentas y transparencia entre actores: gobierno, grupos de interés, sociedad, docentes, etc.

- Dimensión económica: se basa en la relación entre la teoría de capital humano y la creencia de que mejorar las habilidades medidas en evaluaciones internacionales producirá crecimiento económico (Hanushek y Woessmann, 2010). Sobre esta premisa, los resultados de pruebas como PISA se convierten en señales para inversionistas, quienes las interpretan como indicadores de factor humano capacitado. En algunos casos, el condicionamiento de apoyos internacionales a la disponibilidad de evaluaciones es un motor de desarrollo de estos sistemas. Finalmente⁹, y no menor, es la creciente industria evaluativa que se ha desarrollado en torno a la aplicación de estas pruebas y que cabildea por su permanencia y crecimiento.
- Dimensión técnica: la participación en estas evaluaciones se percibe como una oportunidad para el desarrollo de capacidades evaluativas o, en su caso, para suplir una falta de éstas. Asimismo, las reuniones y talleres de preparación de las pruebas brindan a los equipos técnicos y de tomadores de decisiones de distintos países una oportunidad para conversar e intercambiar ideas sobre políticas educativas, planeamiento e investigación, además de que en algunos casos se ha establecido un sistema de tutelaje entre países con más experiencia y países que recién inician o tratan de consolidar sus sistemas de evaluación.
- Dimensión sociocultural: también conocida como “modelos de emulación normativa”, la adopción de políticas educativas, en este caso de evaluación, tiene como propósito legitimar al Estado y presentarlo como moderno, responsable y comprometido frente a la comunidad internacional.

En la realidad, las motivaciones plasmadas en estas dimensiones se entremezclan y pueden cambiar con el tiempo, ya que el proceso de elaboración de políticas, incluyendo las políticas de evaluación, es confuso, poco estructurado e impulsado por muchos factores, no solamente el técnico (Addey y Sellar, 2019; Porter, 2010; Ravela, 2015).

El marco de motivaciones utilizado por Addey y Sellar también resulta válido para explicar la incursión de los países en evaluaciones nacionales a

⁹ Empresas y consultores que ofrecen sus servicios especializados a las agencias encargadas del desarrollo y aplicación de las pruebas estandarizadas

gran escala, ya que las dinámicas políticas, económicas, técnicas y socioculturales que operan a nivel internacional tienen su reflejo hacia el interior de los países¹⁰.

Antes de los años noventa, los esfuerzos de la mayoría de los países en la región estaban enfocados en la cobertura del sistema educativo, entendida como matriculación en los niveles obligatorios¹¹. Sánchez Cerón y Corte Cruz del Sagrario identifican tres rasgos en el contexto regional que marcaron la década de los noventa: la adopción de políticas neoliberales para revertir las crisis económicas, la reestructuración del Estado y apertura de mercados, y una disminución del gasto social favoreciendo la inversión privada (Sánchez Cerón y Corte Cruz del Sagrario, 2013). En el mismo artículo, los autores denuncian la inequidad social que se ha acentuado en las últimas décadas.

Jiménez Moreno (2016) sostiene que fue la apertura de los gobiernos regionales al mercado global, en esa misma década, lo que despertó un interés por la competitividad y enfocó la atención en la calidad de los sistemas educativos, con la subsecuente inversión en sistemas de evaluación para intentar medirla.

Estas últimas interpretaciones encuentran sustento en el papel que jugaron el Banco Mundial, el Banco Interamericano de Desarrollo, la UNESCO y la Agencia de los Estados Unidos para el Desarrollo Internacional (USAID) en el desarrollo de los sistemas de evaluación de Guatemala, Ecuador, Panamá, Honduras y Nicaragua, entre otros; así como el interés en México por participar en evaluaciones internacionales a partir de su incorporación a la OCDE en 1991 (Martínez Rizo, 2008).

En lo que se refiere a la participación de países en la región en pruebas internacionales o regionales, Ferrer y Fiszbein identifican tres motivaciones que se han mantenido desde el año 2003:

¹⁰ Por ejemplo, tanto en México como en Chile, los resultados de evaluaciones fueron pieza clave para generar el sentido de urgencia requerido para aprobar las reformas educativas estructurales de 2011 y 2013, respectivamente (Ayala & Sepúlveda, 2018; Mexicanos Primero, 2012)

¹¹ Esto también implicó cambios en los niveles considerados obligatorios en cada país, lo cual se puede ver en profundidad en: Instituto Nacional para la Evaluación de la Educación (2018). Gaceta de la Política Nacional de Evaluación Educativa, Volumen 12. https://www.inee.edu.mx/wp-content/uploads/2019/01/G12_ESP.pdf.

- El interés por compararse con países “pares” de la región (en el caso del LLECE) y con otros industrializados (en PISA).
- La oportunidad de fortalecimiento técnico del país.
- El impulso por actuar en la mejora de los sistemas educativos que se deriva del debate en torno los resultados (incluso cuando estos son bajos).

Estas motivaciones se encuentran presentes aún en las últimas aplicaciones como PISA para el Desarrollo de 2018, en la que algunos países como Ecuador y Guatemala participaron en gran medida gracias al apoyo financiero que obtuvieron de parte del Banco Interamericano de Desarrollo y por la posibilidad de fortalecer las capacidades técnicas de los equipos nacionales (OECD, 2019). Además, los países reconocen beneficios económicos y políticos de vincularse con la OCDE (Ferrer y Fiszbein, 2015).

Más adelante se verá cómo las distintas motivaciones influyen o moldean el tipo de usos que se ha dado a los resultados de las evaluaciones.

2.2 Primera oleada: aparición de evaluaciones de gran escala y creación de sistemas de evaluación

La creación de los sistemas de evaluación supuso un gran esfuerzo regional de desarrollo de capacidades evaluativas. En la década de los noventa se crearon nuevas instituciones evaluadoras en la región, y se fortalecieron las existentes. Países como Ecuador, Cuba, El Salvador, México, Nicaragua, Panamá y Perú desarrollaron áreas de evaluación dentro de los ministerios de educación; en Bolivia y Colombia se crearon instituciones nuevas; mientras que en Brasil se ampliaron las funciones del Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). Otro grupo de países comenzó la evaluación apoyándose en la experiencia técnica de universidades, formando alianzas con el Ministerio. Tal fue el caso de Chile, Costa Rica, Guatemala y Honduras (Martínez Rizo, 2008).

La consolidación de estos sistemas en los últimos 30 años ha sido dispar debido a circunstancias técnicas y políticas:

“Algunos países han mantenido sus sistemas en funcionamiento con continuidad, aun cuando se produjeran cambios de enfoque o variaciones, a veces accidentadas, en la institucionalidad de los mismos. Otros países

han sufrido importantes discontinuidades y han tenido que comenzar prácticamente desde cero en repetidas oportunidades o tendrán que hacerlo en el futuro próximo¹²” (Ravela, et al., 2007, p. 6)

A nivel técnico, las primeras evaluaciones desarrolladas en la región se centraron en aprendizajes de lenguaje y matemáticas, utilizando la teoría clásica de pruebas y un enfoque normativo que devuelve resultados, en la mayoría de los casos, en forma de puntaje o número de aciertos (Martínez Rizo, 2008).

Durante este tiempo, se vio muy poco impacto de los resultados en las políticas educativas¹³. Frente a una incipiente cultura de evaluación, se dieron casos en los que no se publicaban los resultados bajos por temor al impacto negativo que esto pudiera generar (Ravela, et al., 2007). Un ejemplo de esto fue la aplicación de TIMSS en México en 1995 (Martínez Rizo, 2008). Se dieron experiencias similares en Bolivia, Ecuador, Guatemala y Perú (Ferrer, 2006).

2.3 Segunda oleada: evolución institucional, consideración de factores asociados y devolución de resultados por niveles de logro

Hacia el fin de los noventa y durante los primeros años del siglo XXI, conforme se iban consolidando los sistemas de evaluación, se popularizó el modelo de evaluación de procesos y comenzaron a utilizarse cuestionarios de contexto para acompañar la interpretación de resultados de aprendizaje (Fortín Morales, 2013; Martínez Rizo, 2008). En este sentido, Cuba y Argentina fueron pioneros en la consideración de factores asociados al aprendizaje. También, otros países ampliaron las asignaturas o campos de conocimiento evaluados.

En el año 2000 comenzó una revisión de las teorías psicométricas utilizadas, y para el año 2008 la mayoría de los países en la región, incluyendo a Bolivia, Argentina, Brasil, Chile, Guatemala y Ecuador, utilizaban pruebas en las que los resultados se reportaban por niveles de logro, en vez de

¹² Para una descripción más detallada y actualizada de los sistemas de evaluación de la región, podrá consultarse el fichero sobre las características de los sistemas de evaluación en América Latina.

¹³ Una excepción es Chile, donde el SIMCE se utilizó dentro del sistema de selección de escuelas y asignación de fondos.

puntajes (Martínez Rizo, 2008)¹⁴.

En el marco institucional, los ministerios de educación empezaron a asumir funciones de evaluación previamente compartidas con universidades, y en otros casos se crearon instituciones evaluadoras fuera de los ministerios con mayor o menor grado de autonomía (Martínez Rizo, 2008). Se fortalecieron prácticas de publicación de resultados, aunque con una visión simplista que esperaba ver una mejora inmediata en los resultados de aprendizaje, obviando la complejidad de los procesos y sistemas educativos (Martínez Rizo, 2008). El bajo involucramiento del público en general en las pruebas generaba poca discusión en torno a los resultados una vez que se publicaban (Ferrer, 2006).

2.4 Tercera oleada: desarrollo de mediciones más complejas, incorporación de elementos no cognitivos en la evaluación y preocupación por el uso de los resultados

La última década se ha visto marcada por un esfuerzo para responder a los retos que han surgido a partir de la implementación de los sistemas de evaluación y la evolución de los sistemas educativos: ampliar el concepto de calidad educativa, explorar la medición de habilidades blandas y desincentivar los usos no deseados, entre otros.

Ferrer y Fiszbein (2015) reconocen como importantes innovaciones en la región el Índice de Desenvolvimento da Educação Básica (IDEB), de Brasil, y el Índice Sintético de Calidad Educativa de Colombia (ISCE). Ambos esfuerzos señalan un interés por utilizar los resultados de las evaluaciones con información adicional¹⁵ que permita dar cuenta de un concepto más amplio de calidad educativa, además de dar seguimiento en el tiempo e informar decisiones de política. Una crítica de los autores a estos índices es que utilizan factores asociados simples, en lugar de factores más complejos como los que se recogen en evaluaciones muestrales. Otros países como Chile y Ecuador también han avanzado en la construcción de índices de esta naturaleza.

¹⁴ Bolivia transitó de pruebas normativas a criterios en 2000, Brasil hizo lo mismo en 2001 y Argentina, en 2005. Chile y Guatemala cambiaron la devolución de resultados de puntajes a niveles de logro en 2006 y Ecuador, en 2008, cambió de teoría clásica a respuesta del ítem.

¹⁵ El IDEB utiliza resultados de aprendizaje, tasas de aprobación anual y tasas de abandono; el ISCE utiliza rendimiento escolar, aprobación, abandono y clima escolar.

De acuerdo con la tendencia internacional reciente de reivindicar el valor de habilidades y conocimientos no académicos para el desarrollo personal y profesional (Kautz, Heckman, Diris, ter Weel, y Borghans, 2014), algunos países de la región han incorporado, en la última década, elementos de evaluación de aprendizajes no cognitivos. De igual manera se ha reforzado y ha aumentado la aplicación de cuestionarios de contexto dirigidos a estudiantes, docentes y familias.

La Agencia Calidad de la Educación (ACE), en Chile, ha incorporado la evaluación Indicadores de Desarrollo Personal y Social (IDPS): autoestima y motivación escolar, clima y convivencia escolar, hábitos de vida saludable, y formación y participación ciudadana (Ayala y Sepúlveda, 2018).

En México el constructo equivalente se ha denominado “habilidades socioafectivas” y se refiere a aprendizajes de saberes que favorecen la convivencia escolar (INEE, 2018).

Costa Rica, Ecuador y República Dominicana también incluyen medidas de habilidades socioemocionales en alguno de sus niveles educativos.

Aunque los marcos conceptuales para la medición de estos constructos, así como los mecanismos de evaluación, difieren entre países y aún están en desarrollo, es de esperarse que en los próximos años se refuercen y amplíen a más países de la región, mediante la incorporación de evaluaciones nacionales o la participación en evaluaciones internacionales de gran escala.

Conforme se ha ido evidenciando la limitación formativa de las evaluaciones estandarizadas¹⁶, han surgido otras alternas, como parte de un mismo sistema de evaluación, para proveer de información a distintos niveles del sistema educativo. En el caso de Chile, Ayala y Sepúlveda describen los modelos de evaluaciones formativas y progresivas de la ACE:

Ambas—sobre todo la segunda— son para un uso interno en el colegio: no arrojan resultados públicos, sino que están abocadas a monitorear los aprendizajes a lo largo del año (Ayala y Sepúlveda, 2018, p. 3).

Como se abordará más adelante, el tema del uso de los resultados es un punto de interés y controversia en los países de la región. La asociación

¹⁶ Ver apartado sobre factores que influyen en el uso de resultados y las posturas frente a las evaluaciones estandarizadas.

de esquemas de incentivos a centros escolares o maestros, así como las prácticas de ranking y una preocupación por los efectos homogeneizadores de las evaluaciones estandarizadas, han provocado prácticas no deseadas y posiciones encontradas en torno a las evaluaciones de gran escala (Jiménez Moreno, 2016; Loza, 2018; Sánchez Cerón y Corte Cruz del Sagrario, 2013).

2.5 Marcos institucionales de las unidades evaluadoras

El marco institucional en el que se aplican las evaluaciones estandarizadas es un elemento que influye en el uso de sus resultados desde múltiples frentes: la estabilidad de los equipos técnicos a cargo del diseño y aplicación de las pruebas, la asignación y manejo de recursos humanos y presupuestales, los canales de comunicación entre evaluadores y usuarios de la información y el manejo político del discurso en torno a las evaluaciones, entre otros.

Las instituciones evaluadoras de la región se pueden clasificar en dos tipos, de acuerdo con el marco institucional que las rige: instituciones autónomas (o semiautónomas) y áreas o unidades dentro de los ministerios de educación¹⁷.

El informe del Banco Mundial titulado *¿Qué ha sucedido con los sistemas de evaluación de aprendizajes en América Latina?* Lecciones de la última década de experiencia presenta una actualización del estado de los sistemas de evaluación en sus marcos institucionales, en la difusión y los usos de la información, y en la participación en pruebas internacionales, enfocándose en cinco países de la región: Brasil, Colombia, Guatemala, México y Perú (Ferrer y Fiszbein, 2015).

De los países considerados en el informe, México, Colombia y Perú han tenido cambios recientes en su andamiaje institucional. En 2009 se añadieron atribuciones al Instituto Colombiano para la Evaluación de la Educación (ICFES), para “ofrecer servicios de evaluación, recaudar ingresos y retener excedentes para ser reinvertidos en mejoras técnicas y desarrollo de programas” (Ferrer y Fiszbein, 2015, p. 5). Por su parte, la Dirección General de Evaluación, Investigación y Estándares Educativos de Guatemala fue

¹⁷ Para una clasificación pormenorizada de los marcos institucionales en cada país, se podrá consultar el fichero de características de los sistemas de evaluación.

constituida legalmente en 2007, aunque llevaba varios años en operación. En México, el Instituto Nacional para la Evaluación de la Educación (INEE) pasó de ser un organismo descentralizado a ser un ente autónomo en 2013 (LINEE, 2013). Sin embargo, esta autonomía fue revocada en una reforma constitucional de mayo de 2019, para regresar a ser un organismo descentralizado con una identidad nueva (CPEUM, 2019).

En 2011, Chile y Ecuador tuvieron también reformas importantes en el tema, con la descentralización de la ACE (Ley 20529, 2011) y la creación del Instituto Nacional de Evaluación Educativa (INEVAL) como un organismo autónomo (LOEI, 2011), respectivamente.

Si bien en los últimos años se percibe una tendencia a separar las instancias evaluadoras de los ministerios de educación y dotarlas de mayores atribuciones, actualmente existe un debate al respecto que se abordará más adelante.

Tabla 4. Clasificación de instituciones evaluadoras, a nivel nacional, por tipo de relación con la autoridad educativa u otras instancias.

RELACIÓN CON AUTORIDAD EDUCATIVA	PAÍS	INSTITUCIÓN EVALUADORAS
SON UNIDADES DENTRO DEL MINISTERIO DE EDUCACIÓN	Argentina	Ministerio de Educación, Cultura, Ciencia y Tecnología, Secretaría de Evaluación Educativa
	Costa Rica	Ministerio de Educación Pública, Consejo Superior de Educación, Dirección de Gestión y Evaluación de la Calidad
	El Salvador	Dirección de Evaluación de los Aprendizajes del Ministerio de Educación

<p>TIENEN CIERTO GRADO DE AUTONOMÍA, PERO COMPARTEN RESPONSABILIDADES DE EVALUACIÓN CON EL MINISTERIO DE EDUCACIÓN U OTRAS INSTANCIAS</p>	Guatemala	Dirección General de Evaluación e Investigación Educativa (DIGEDUCA) del Ministerio de Educación
	Panamá	Ministerio de Educación (MEDUCA)
	Paraguay	Instituto Nacional de Evaluación Educativa (INEE) ¹⁸ Dirección General de Planificación Educativa del Ministerio de Educación y Cultura
	Honduras	Proyecto Mejorando el Impacto al Desempeño Estudiantil en Honduras (Proyecto MIDEH) Dirección General de Currículo y Evaluación de la Secretaría de Educación
	México	Instituto Nacional para la Evaluación de la Educación (INEE) ¹⁹ Secretaría de Educación Pública
	Perú	Ministerio de Educación del Perú (MINEDU) Oficina de Medición de la Calidad de los Aprendizajes (UMC) Instituto Nacional de Estadística e Informática (INEI)
	República Dominicana ²⁰	Instituto Dominicano de Evaluación e Investigación de la Calidad Educativa Dirección de Evaluación de la Calidad del Ministerio de Educación

¹⁸ Creado en la Ley N° 5749 “Que establece la Carta Orgánica del Ministerio de Educación y Ciencias” de noviembre de 2018.

¹⁹ Hasta el 16 de mayo de 2019.

²⁰ La evaluación de primera infancia ENHOGAR-MICS es realizada por la Oficina Nacional de Estadísticas.

3. Marco conceptual sobre usos de resultados de evaluaciones

En este apartado se presenta una síntesis de referentes conceptuales en torno al uso de los resultados de las evaluaciones con el fin de enmarcar los apartados siguientes. Los primeros modelos se centran en tipificar los usos de las evaluaciones. Los últimos, abordan el tema de las teorías de cambio relacionadas con el uso de las evaluaciones y la mejora educativa

3.1. Tipología de usos de evidencia

Carol Weiss propone una tipología de usos de evidencia científica, en general, orientada a clarificar este concepto de acuerdo con el rol que la información juega en una situación específica. Esta clasificación fue evolucionando entre 1988 (Weiss, 1979) y 1998 (Weiss, 1998) para concretarse de la siguiente forma:

- **Uso instrumental:** los resultados influyen en la toma de decisiones sobre un programa o política. Su uso implica tres condiciones: los hallazgos son poco controversiales sin enfrentamiento de intereses, los cambios sugeridos son relativamente sencillos de implementar, el entorno de implementación permanece relativamente estable. En casos de parálisis o emergencia, también se puede recurrir a los resultados para informar decisiones.
- **Uso conceptual:** aún sin llegar a hacer cambios, los involucrados modifican y enriquecen su concepción de la situación o programa. Se identifican fuerzas y debilidades que podrán considerar cuando se den las condiciones para actuar.
- **Uso persuasivo:** los resultados se utilizan para sustentar decisiones previamente tomadas que requieren legitimidad.
- **Uso “ilustrado”:** se refiere a los casos en que los resultados influyen en eventos e instituciones que exceden el alcance de la evaluación. Por ejemplo, las revisiones de literatura o síntesis de conclusiones sobre estudios de un mismo tema o programa que generan conversaciones entre académicos, políticos, organismos de cabildeo, etc. y provocan cambios en la agenda y los paradigmas de política, así como en las opiniones y

creencias en comunidades.

En la literatura de la región frecuentemente se ven alusiones a esta clasificación, sobre todo cuando hay referencias de uso de tipo instrumental y persuasivo o simbólico.

Silva, D'Angelo, Fernández, Martínez y Rodríguez (Silva et al., 2012) clasifican de manera más concreta la forma en que los actores educativos del sistema, a nivel macro, acceden y utilizan los resultados de las evaluaciones, a partir de un estudio realizado con autoridades educativas en México:

- No uso: por desconocimiento, falta de oportunidad en la devolución de resultados o poca pertinencia debido a niveles de desagregación muy bajos o aspectos coyunturales de operación o políticos.
- Uso mediatizado: se hacen solicitudes de información puntuales sin indagar de forma directa las fuentes, sino a través de terceros dentro o fuera de la organización²¹.
- Uso consultivo: de profundidad variable, se accede directamente a las fuentes sea para buscar, corroborar o contrastar un dato puntual o para un ejercicio más amplio de diagnóstico o planeación. Hay un mayor conocimiento de las fuentes de información²².
- Uso integrado en acciones, programas o políticas: el uso de los resultados se orienta a sustentar acciones o alimentar el diseño de políticas específicas.

Si bien la nomenclatura de Weiss permite identificar la intención de uso de la información, esta clasificación resulta útil para diferenciar con mayor detalle el comportamiento de distintos actores en torno al manejo de los resultados, desde un enfoque más operativo.

3.2 Propósitos de la evaluación

En el libro *Para comprender las evaluaciones educativas* (2005), Ravela describe dos grandes clasificaciones de evaluaciones educativas, de acuerdo con el tipo de consecuencias o propósito que tengan:

Aquellas que tienen consecuencias directas importantes para individuos

²¹ Fue el uso predominante entre las autoridades educativas de más alto mando entrevistadas en el estudio.

²² Uso observado, principalmente, entre niveles de mando medio y cuerpos técnicos, así como equipos de asesores, que participaron en ese estudio.

o instituciones –los exámenes para aprobar un curso, las pruebas de selección, las evaluaciones que definen una calificación para los maestros o determinan premios en dinero– y aquellas que tienen como propósito principal aprender para mejorar, pero que no tienen consecuencias “fuertes” (p. 28).

A partir de esta agrupación, el autor desarrolla la siguiente clasificación más pormenorizada:

- Acreditación o certificación.
- Ordenamiento para selección.
- Toma de decisiones para la mejora (“blandas”).
- Toma de decisiones “duras” (con consecuencias de continuidad, ampliación, etc.).
- Establecimiento de incentivos y sanciones.
- Rendición de cuentas y responsabilización de resultados.

De acuerdo con esta clasificación, las evaluaciones realizadas a alumnos pueden utilizarse para acreditación o certificación, ordenamiento para selección, toma de decisiones “blandas” y establecimiento de incentivos.

Las evaluaciones de docentes y de centros educativos comparten la misma variedad de fines²³, aunque si se consideran los resultados de los alumnos, puede darse el caso en que una evaluación de gran escala tenga consecuencias “blandas” para los alumnos, pero de ordenamiento o incentivos para docentes y escuelas.

3.3 Interacción entre visiones y funciones de la evaluación y el uso de resultados

Fortín Morales (2013) hace un análisis de la relación que existe entre la visión y función que se espera de la evaluación, y el diseño y difusión de los resultados, a partir de las narrativas de actores educativos de Guatemala en torno a las evaluaciones estandarizadas. Si bien el autor advierte que no se trata de un modelo teórico, la clasificación que presenta resulta útil para analizar la realidad en otros países de la región.

²³ A los centros se añade el propósito de rendición de cuentas y responsabilización.

El autor identifica tres visiones sobre la evaluación en los distintos actores involucrados, tanto en el diseño como en el uso de las evaluaciones:

- Como herramienta para la gestión: la evaluación es un elemento más para la gestión del sistema como un todo. Es un componente del aseguramiento de la calidad. La relación que entabla con los actores escolares va enfocada a mejorar el servicio que prestan.
- Como objeto de análisis psicométrico: el esfuerzo se centra en producir información técnicamente sólida y válida que pueda ser útil a tomadores de decisiones. Opera sobre el supuesto de que los tomadores de decisiones tienen preguntas que necesitan responderse, y que estas coinciden con las que se plantean en las evaluaciones.
- Como fuente de información pedagógica: el interés está en relacionar la evaluación con lo que hacen los docentes y estudiantes, más que ver el sistema educativo como un todo.

En torno a estas visiones, surgen en la narrativa cuatro funciones esperadas de los resultados que, en ocasiones, se entremezclan:

- Control de calidad: con un enfoque más sumativo y de categorías bien definidas, dan cuenta del cumplimiento o no de resultados esperados.

		VISIÓN DE LA EVALUACIÓN			Comunicación de resultados
		Herramienta de gestión	Objeto de análisis psicométrico	Fuente de información pedagógica	
FUNCIÓN ATRIBUIDA A LOS RESULTADOS	Para control de calidad	*	+		Ranking Certificación Acreditación
	Como ejercicio formativo	+		*	Dependerá del actor evaluado
	Como investigación	+	*		Reportes técnicos Reportes de valor agregado Reportes de factores asociados
	Para brindar orientación pedagógica	+	+	*	Sugerencias pedagógicas

* Uso prioritario de acuerdo con la visión.

+ Uso admitido, pero no intencionado o dirigido por la visión.

Adaptado de "Evaluación educativa estandarizada en Guatemala: Un camino recorrido, un camino por recorrer" (Fortín Morales, 2013).

Aún sin referirse intencionadamente a esta clasificación, muchos de los trabajos analizados en los apartados siguientes reflejan estas visiones de la evaluación y los usos que se plantean corresponden a las funciones aquí descritas.

3.4 Teorías de cambio de las evaluaciones en gran escala

En el artículo “El uso de los resultados de las evaluaciones en la mejora de los procesos de enseñanza” (Ravela, 2015), el autor identifica seis lógicas de cambio presentes en distintos sistemas de evaluación de la región, que informan su diseño y operación:

- Lógica del mercado educativo: las evaluaciones censales y con resultados públicos generan competencia entre las escuelas que fomenta procesos de mejora.
- Lógica de incentivos económicos directos: el otorgamiento de incentivos económicos a partir de los resultados de evaluaciones censales estimula a los docentes a mejorar su práctica, contribuyendo a mejores aprendizajes por parte de los alumnos.
- Lógica de presión social sobre las escuelas: la publicación de resultados de evaluaciones genera una demanda de resultados por parte de las familias y la sociedad y una responsabilización de los centros escolares.
- Lógica de desarrollo profesional: las evaluaciones estandarizadas voluntarias con resultados no públicos invitan a los docentes a revisar sus prácticas y el proceso de aprendizaje de los alumnos, sin la presión de otras consecuencias que podrían distraer el propósito de mejora continua.
- Lógica tecnocéntrica: las evaluaciones, junto con información del contexto y escolar, ofrecen información valiosa para la toma de decisiones en política educativa al identificar los factores con mayor incidencia sobre los resultados.
- Lógica de información estadística: los resultados de las evaluaciones colocan el tema del aprendizaje en la agenda pública y de la ciudadanía sin pretender provocar cambios directamente.

Por su parte, Fortín Morales (2013) describe cuatro lógicas distintas de impacto de las evaluaciones en tres niveles del sistema educativo, de acuerdo

con la función que se les atribuye. Estas lógicas de alguna manera se incluyen y coinciden con las descritas por Ravela.

Niveles de impacto:

1. Sistema: las estructuras administrativas y las acciones que implementan sus instituciones.
2. Política: directrices y lineamientos de operación.
3. Aula: acciones en el aula como parte del proceso de enseñanza y aprendizaje.

Tabla 6. Aporte de la evaluación concebida como herramienta para implementación del control de calidad

SISTEMA	EVALUACIÓN → JUICIO → ACCIÓN
POLÍTICA	
AULA	

Tabla 7. Aporte de la evaluación concebida como ejercicio formativo

SISTEMA	EVALUACIÓN → ACCIÓN
POLÍTICA	
AULA	EVALUACIÓN → ACCIÓN

Tabla 8. Aporte de la evaluación concebida como investigación

	INFORMACIÓN ADICIONAL ↓
SISTEMA	EVALUACIÓN → FACTORES ASOCIADOS
POLÍTICA	↘ ACCIÓN
AULA	↘ ACCIÓN

Tabla 9. Aporte de la evaluación concebida como fuente de orientación pedagógica

SISTEMA	EVALUACIÓN → INTERPRETACIÓN → COMUNICACIÓN
POLÍTICA	
AULA	↓ ACCIÓN

Tablas 6-9 Adaptadas de "Evaluación estandarizada en Guatemala: Un camino recorrido, un camino por recorrer" (Fortín Morales, 2013).

En los siguientes apartados se mencionarán ejemplos de la aplicación de estas lógicas de cambio y algunos de los resultados que han provocado.

4. Uso de resultados de evaluaciones en políticas educativas e intervenciones en la región

Si bien los sistemas de evaluación en la región surgieron de un interés por medir la calidad de la educación, las lógicas de impacto para mejorar a partir de esas mediciones, y por lo tanto los usos, son muy variados. Por ejemplo, Chile adoptó un enfoque de control de calidad mediante lógica de mercado, cargando de peso a los resultados del Sistema de Medición de la Calidad de la Educación (SIMCE) como indicador para la elección de escuela por parte de la familia (Ravela, 2015). En Colombia, se entregan incentivos a escuelas y docentes con base en los avances en el Indicador Sintético de Calidad que incluye los puntajes de la prueba SABER para fomentar la mejora de su práctica y los aprendizajes de los alumnos (Cabal, Milena y Castro, 2018). En otros países, como Honduras, la evaluación tanto de docentes como de alumnos no tiene consecuencias y se aspira a que sirva de insumo para la responsabilización y la toma de decisiones (FEREMA, 2017).

Entre los usos que se identifican en la literatura de la región se recogen ejemplos de cómo se han utilizado los resultados de evaluaciones para:

- Definir la agenda política.

- Informar ejercicios de planeación estratégica a distintos niveles del sistema.
- Focalizar programas e intervenciones.
- Orientar la definición del currículo y materiales.
- Evaluar o incentivar a docentes.
- Clasificar o generar un ranking de escuelas.

4.1 Definición de la agenda política

Refiriéndose al discurso político y la agenda educativa, Pedro Ravela señala que “es notorio que la construcción y difusión de información sobre aprendizajes han contribuido significativamente a colocar estos últimos como eje de preocupaciones y debates, algo que no ocurría antes de los años noventa” (Ravela, 2015). Rivas (2015), en su análisis de países en la región a quince años de participación en PISA, reconoce que los sistemas de evaluación a gran escala pusieron sobre la mesa el tema de equidad en la educación al visualizar las brechas de logro existentes entre diversos segmentos de la población.

Poco a poco, se ha ido instalando en la región una cultura de evaluación o de políticas basadas en evidencia, al menos a nivel de discurso, en la que se reconoce la necesidad de tener evaluaciones para sustentar el proceso de toma de decisiones en la política educativa (Ferrer y Fiszbein, 2015; Jiménez Moreno, 2016).

La formalización de los usos de los resultados es muy variada, desde su utilización retórica en discursos y como justificación de decisiones previamente tomadas, hasta intervenciones sobre áreas específicas del sistema.

El ejemplo de PISA ilustra este punto con claridad. La cobertura mediática y retórica del estudio suele utilizarse para impulsar decisiones o acciones del gobierno ya acordadas (Addey y Sellar, 2019; Jiménez Moreno, 2016; Mosqueira y Robinson, 2018). Por otro lado, a partir de los resultados de PISA, se han dado en la región ejercicios de revisión y replanteamiento de lo que los alumnos estaban aprendiendo. De ahí se derivaron proyectos y programas enfocados en revisar los currículos nacionales, el replanteamiento de la formación docente y la asignación de nuevos roles y responsabilidades al director (Rivas, 2015).

En México, los resultados de ENLACE y PISA se utilizaron en algunos estados para elaborar perfiles de egreso de estudiantes y alinear la política evaluativa estatal a los estándares federales. Sin embargo, en otras ocasiones, se encontró que los resultados se utilizan en “diagnósticos” que realmente son justificaciones (Silva, D’Angelo, Fernández, Martínez, y Rodríguez, 2012). Por su parte, Loza (2018) propone que la interpretación contextualizada de los resultados de pie a la implementación de políticas diferenciadas.

4.2 Planeación estratégica

En lo que se refiere al uso de resultados en procesos de planeación estratégica, el ICFES de Colombia realiza talleres con las Secretarías de Educación de los departamentos para orientar en el uso e interpretación de los resultados y utilizarlos como insumo para el diseño de estrategias de los planes de mejora, incluyendo las etapas de diagnóstico, diseño de acciones y seguimiento (Mineducación, 2015).

En México, Silva et al. (2012) documentaron casos de uso de resultados de evaluaciones para la asignación de presupuesto a un nivel o sector específico, y como referencia o datos adicionales dentro de diagnósticos sectoriales o programáticos más amplios. En Chile, la Comisión para el Desarrollo y Uso del SIMCE, reportó un uso amplio de los resultados de SIMCE en el diseño y evaluación de políticas, como indicadores de metas nacionales, así como instrumentos de evaluación de programas e intervenciones, tanto a nivel nacional como de escuelas (Sandoval-Hernández, 2016).

4.3 Focalización de programas e intervenciones

Existe menos información sobre el uso concreto de los resultados de evaluaciones en programas e intervenciones específicas, probablemente porque esta información se encuentra en documentos internos o reglamentos de programas. Se recuperan algunas experiencias concretas de México en el que se utilizaron resultados de ENLACE como criterio de participación e instrumento de rendición de cuentas del programa Escuelas de Calidad (Silva et al., 2012). De igual forma, los lineamientos para el Sistema de Asesoría y Acompañamiento a las Escuelas (SISAAE) refieren a los

resultados de evaluaciones externas e internas para identificar las necesidades pedagógicas y definir los objetivos y metas para el ciclo escolar (CNSP, 2019).

En Chile, el propio Ministerio de Educación señala que los resultados de las evaluaciones se pueden utilizar para focalizar establecimientos que pueden beneficiarse de intervenciones, apoyos o programas como 900 Escuelas, o recibir fondos para proyectos de mejora (Sandoval-Hernández, 2016; Rivas, 2015). En Uruguay, se utilizaron los resultados para mantener y expandir el programa de escuelas de tiempo completo, y en Bogotá para tomar decisiones sobre inversiones de infraestructura y comidas escolares (Rivas, 2015).

4.4 Orientación del currículo y materiales

En México, se registró un caso a nivel federal de uso de los resultados de los aprendizajes para orientar la decisión sobre los temas a reforzar en los materiales educativos, ajustes al currículo y diseño de cursos (Silva et al., 2012). Bogotá, Paraná y São Paulo²⁴ incluyen en sus sistemas de evaluación docente un componente de formación continua que aborda cómo mejorar el desempeño de los alumnos en pruebas estandarizadas (Rivas, 2015). En Chile, las pruebas estandarizadas han fomentado un apego al currículo establecido (Sandoval-Hernández, 2016) y, de manera general, se ha visto una influencia de PISA en la definición de estándares de aprendizaje en la región, aunque también esta influencia se enfrenta a otras dinámicas internas de los países, como presiones sindicales por mantener puestos de trabajo o demandas sociales para priorizar ciertos saberes (Rivas, 2015).

4.5 Incentivos docentes

De acuerdo con el estudio del PREAL, los resultados de pruebas estandarizadas se interpretan como “un indicador de la calidad del trabajo de los docentes y un criterio para otorgar incentivos económicos” (Ravela, et al., 2007, p. 14). Algunos países, como Colombia y México²⁵, han establecido incentivos ligados al resultado de evaluaciones a través de programas como

²⁴ Paraná y Sao Paulo operan sistemas de evaluación subnacionales.

²⁵ En México, la práctica se suspendió en 2013 con la sustitución de ENLACE con el Plan Nacional de Evaluación de Aprendizajes (PLANEA).

Incentivos a la Calidad Docente, en el primer caso, y a los Estímulos de Calidad Docente, en el segundo (Silva et al., 2012; Cabal, Milena, y Castro, 2018). De manera informal, en Chile, se reportaron pagos a docentes cuyos alumnos obtenían los mejores puntajes, a pesar de las advertencias de la propia Agencia de la Calidad de la Educación en contra de estas prácticas (Sandoval-Hernández, 2016).

4.6 Clasificación o ranking de escuelas

Finalmente, se debe mencionar el uso de la clasificación o ranking de escuelas para los siguientes fines:

- i) fomentar la responsabilidad de las escuelas por sus resultados; ii) entregar a las familias información para la toma de decisiones sobre la escuela a la cual enviar a sus hijos; iii) promover la competencia entre los centros educativos por obtener mejores resultados.

Utilizar los resultados de evaluaciones estandarizadas para entregar incentivos económicos a las escuelas que obtienen mejores resultados o mejoras en relación a sus resultados en evaluaciones anteriores (Ravela, et al., 2007, p. 14).

Tal es el caso de Chile, como se mencionó al inicio, aunque en años recientes se ha intentado alejar de la práctica de ranking para enfatizar el uso formativo de las evaluaciones (Ayala y Sepúlveda, 2018; EDUY21, sf). En Guatemala, se devuelven resultados por escuelas y se ordenan según resultados de alumnos, considerando el nivel socioeconómico (Martínez Rizo, 2008). Si bien en México no hay consecuencias “duras” para las escuelas, la publicación de resultados y rankings elaborados por los medios de comunicación o las propias autoridades estatales genera presión sobre los supervisores, directores y equipos docentes, o deriva en prácticas de simulación (Sánchez Cerón y Corte Cruz del Sagrario, 2013). En el caso de Honduras no se emiten rankings como tal, sino que se publican los resultados desagregados a nivel municipal para fomentar la responsabilización y la sana competencia entre departamentos y municipios, con resultados encontrados de legitimación y daño a los centros educativos (FEREMA, 2017). Uruguay fue el primer país en publicar resultados estableciendo comparaciones entre escuelas de nivel socioeconómico similar, y en Brasil se compara a cada

escuela con un grupo de escuelas anónimas que comparten características similares, de acuerdo con el IDEB (EDUY21, sf; Rivas, 2015).

5. Factores que influyen en el uso (o no uso) de resultados de evaluaciones

En este apartado se incluyen trabajos que han documentado factores que influyen en los usos de los resultados de las evaluaciones mencionados anteriormente. Se procuró que, en la medida de lo posible, las alusiones se refirieran específicamente a uso de resultados de evaluaciones de gran escala, diferenciándolos de indicadores u otra información resultante de procesos de evaluación. Existen varios estudios que recuperan información de manera general sobre estos factores de uso de resultados, pero se encontraron pocos trabajos que describieran en detalle, a nivel local o nacional, los matices y dinámicas del uso de los resultados por parte de distintos actores²⁶.

Si bien el uso de resultados, en la realidad, depende de la combinación de múltiples factores que no siempre es posible separar, para efectos de análisis, se agruparon las referencias en cuatro categorías:

- Factores políticos: se refieren a realidades del plano político-gubernamental, así como a la interacción entre las políticas del sistema educativo y sus actores.
- Elementos institucionales: aquellos que se refieren al andamiaje institucional en el que se realizan y usan las evaluaciones.
- Aspectos técnicos: relacionados, ya sea con elementos técnicos de las propias pruebas que influyen en los usos, como las capacidades técnicas para utilizarlos.
- Comunicación y difusión: se refieren a distintas prácticas de reporte y difusión de los resultados.

5.1 Factores políticos

De acuerdo con el análisis de Rivas (2015), durante los primeros quince

²⁶ Se identificó un estudio realizado en 2012 en México y otro en 2013 en Guatemala. Axel Rivas recoge ejemplos en su libro *América Latina después de PISA: Lecciones aprendidas de la educación en siete países (2000-2015)*.

años del nuevo milenio se vio una tendencia regional en renovar la gobernabilidad centralizada de los sistemas educativos. En este marco, las evaluaciones con enfoque de medición de calidad se convirtieron en una herramienta de gestión por resultados y de legitimación de la transparencia de los Estados y gobiernos.

Un factor relacionado con esta legitimidad, que ha cobrado peso en el escenario político en la región, es el tipo de cobertura mediática que se da a los resultados de las evaluaciones. La publicación de los resultados de evaluaciones constituyó un gran paso en materia de transparencia y rendición de cuentas de los gobiernos, como lo reconoce el estudio del PREAL:

Parece haberse superado la etapa en la que las autoridades en varios países interferían en la divulgación de los resultados de las evaluaciones cuando éstos no les parecían favorables. En general, existe en la región una mayor transparencia en la difusión de los resultados (Ravela, et al., 2007, p. 6).

Sin embargo, el inicio de la publicación de resultados de evaluaciones tuvo complicaciones. Se caracterizó por una cobertura en medios sensacionalista, alarmista y descalificadora de los gobiernos y sistemas educativos (Ferrer y Fiszbein, 2015), a pesar de los esfuerzos de las instancias evaluadoras de desechar esas prácticas.

Aun cuando la propia OCDE recomienda contextualizar los resultados de estas pruebas en las condiciones socioeconómicas y culturales de cada país para evitar interpretaciones reduccionistas (OCDE, s/f), es claro que los listados en los que se clasifican los países participantes (ranking) constituyen actualmente el motor del efecto mediático que ha adquirido en casi todos los países, sea para cuestionar el mal desempeño de sus respectivos sistemas educativos o para exaltar sus logros (Jiménez Moreno, 2016, p. 6).

Varios autores coinciden en que la interpretación de los resultados de las evaluaciones es simplista, no ahonda en el enfoque de lo que miden las pruebas, ni considera las dinámicas y complejidades del sistema (Jiménez, 2017; Ravela, et al., 2007)

Con todo, el informe del Banco Mundial (Ferrer y Fiszbein, 2015) reconoce avances en la educación de los medios de comunicación respecto al uso

de resultados:

Los medios de comunicación y los periodistas están cada vez más informados en materia educativa; por lo tanto, ahora pueden ir más allá de los titulares sensacionalistas que eran tan frecuentes y populares en el pasado. Las malas noticias todavía son material para titulares negativos y para provocar duras críticas a las políticas del gobierno, pero al mismo tiempo parecen contribuir a que existan más debates sobre la calidad de la educación en una variedad de foros (Ferrer y Fiszbein, 2015, p. 12).

El informe del Banco Mundial postula que “probablemente, como resultado de esta cobertura masiva, los políticos están poniendo más atención a la importancia y al protagonismo de los resultados de las evaluaciones de aprendizajes, aunque no se sientan necesariamente responsables de esos resultados” (Ferrer y Fiszbein, 2015, p. 12). Sin embargo, lamentan que los políticos sucumben a la presión y “tienden a tomar decisiones apresuradas con el fin de demostrar que están tomando medidas para cambiar la situación” (Ferrer y Fiszbein, 2015, p. 13). En ocasiones, estas decisiones llevan a programas que no han considerado con suficiencia las dinámicas y procesos del propio sistema educativo y su impacto sobre los procesos de enseñanza y aprendizaje. Por lo mismo, se convierten en atajos que no dan los resultados esperados (Ravela, et al., 2007; Rivas, 2015).

Asimismo, resulta difícil reorientar discursos establecidos en los medios, fruto de usos históricos de resultados, para promover un uso de las evaluaciones sin enfoque punitivo o competitivo (Ayala y Sepúlveda, 2018; Ravela, 2015).

Harvey Sánchez (2018) señala que, en ocasiones, los resultados son descartados o utilizados parcialmente, por intereses ajenos a lo educativo. Ferrer (2006) reconoce lo mismo al incluir la voluntad política como uno de los factores que influye sobre el uso de resultados en la región. Como ejemplo se puede mencionar el caso de Bolivia, que en el año 2000 discontinuó la aplicación de la prueba SIMECAL y dejó de participar en las pruebas del LLECE por motivos ideológicos y políticos. Hasta la fecha, el país no ha retomado un sistema de evaluaciones de gran escala, aunque se reincorporó al LLECE para la aplicación del ERCE 2019 (Urquiola, 2011; Addey y Sellar, 2019).

Tanto Addey (2019) como Mosqueira y Robinson (2018) abordan el caso contrario, en el que los resultados de evaluaciones sirven de palanca o justificación para generar un sentido de urgencia, llegando incluso a crear realidades que facilitan a los gobiernos tomar decisiones difíciles o controvertidas.

Desde el punto de vista de la política educativa, refiriéndose al caso de las evaluaciones en Uruguay, Ravela (2015) plantea la probabilidad de que las evaluaciones voluntarias en línea que tienen un alto potencial pedagógico sean utilizadas principalmente por docentes comprometidos y motivados. En consonancia, Rivas (2015) remite a su experiencia e invita a estudiar el fenómeno por el que, en entornos de bajas consecuencias o consecuencias “blandas”, los docentes no hacen el esfuerzo por entender los resultados de evaluaciones. En ambos casos se argumenta que la iniciativa propia del docente no es suficiente para instalar una cultura de uso de resultados en el sistema.

5.2 Elementos institucionales

Como se estableció en el primer apartado del documento, existen varios modelos institucionales en la región para gestionar los sistemas de evaluación. Sánchez (2018) defiende que las instituciones evaluadoras deben tener autonomía e independencia que los alejen de las presiones de poderes externos. Aunque se reconocen las ventajas de objetividad y estabilidad técnica que brinda el tener instituciones autónomas independientes de los ministerios de educación, se ha visto que esta separación dificulta el flujo de información, la vinculación y la apropiación de los resultados por parte de los usuarios (Fortín Morales, 2013; Ferrer y Fiszbein, 2015). Dicho esto, Ferrer y Fiszbein no encontraron diferencias sustanciales al respecto en países donde las unidades de evaluación son parte del ministerio.

Esto sugiere que, más allá de las relaciones institucionales, el peso de las relaciones informales y personales juega un papel importante en el uso de los resultados por parte de las autoridades educativas. Al respecto, Silva et al. (2012) encontraron lo siguiente:

Para el caso de autoridades federales dos condiciones impactaron en los usos predominantes. Por un lado, las particularidades que delimitan

a las competencias de cada dirección general y, por otro lado, la cercanía y trato directo entre esas autoridades y el propio INEE (2012, p. 55).

Los sistemas de evaluación se deben concebir como redes y definir con claridad cómo interactúan dentro del sistema educativo en general (Fortín Morales, 2013). Estas relaciones no siempre son claras, independientemente de si la agencia evaluadora está dentro o fuera del ministerio (Fortín Morales, 2013; Ferrer y Fiszbein, 2015).

Hemos pensado en la evaluación en forma excesivamente autocontenida, no como parte de un conjunto más amplio de acciones y elementos de la política educativa (...) Hemos pensado la evaluación de aprendizajes a gran escala en forma independiente de las evaluaciones de aprendizajes que ocurren diariamente en las aulas y, en muchos casos, de los cambios curriculares. Necesitamos pensar la evaluación como parte de y articulada con un conjunto de acciones propias de la política educativa (Ravela, 2015, p. 80).

Esta falta de articulación frecuentemente pesa sobre el grado de apropiación de la información por parte de los usuarios. En estos casos, la barrera entre los factores políticos y los institucionales no es clara, sobre todo cuando existen conflictos de prevalencia de un área o institución sobre otra, o se percibe rivalidad en la información reportada. Se requiere una mayor documentación de estas situaciones, pero Silva et al. (2012) encontraron informantes en la educación secundaria alta que manifestaban cierto recelo al no ser consultados antes de la publicación de cifras que discrepaban de sus propios datos. En la misma línea, el estudio de Ferrer (2006) señala que la participación y el involucramiento de los usuarios desde las etapas de diseño de las evaluaciones pueden favorecer esta apropiación y consecuente uso de los resultados. Fortín Morales (2013) insta a considerar explícitamente y con mayor claridad los marcos operativos del sistema educativo al momento de pensar la evaluación.

Otro efecto de la falta de claridad en la articulación entre los sistemas de evaluación y otras áreas del sistema educativo es la dificultad para asignar responsabilidades sobre las acciones de mejora. O bien las recomendaciones que se emiten a partir de la evaluación no son vinculantes, o no hay claridad sobre los factores que subyacen a los resultados y a quién compete

emprender las acciones de mejora correspondientes (Silva et al., 2012; Fortín Morales, 2013). Varios autores recomiendan e identifican áreas del sistema educativo que deberían articularse en torno al diseño de políticas, con base en las carencias reveladas por las pruebas de aprendizaje; por ejemplo: evaluación, desarrollo curricular, formación inicial y desarrollo profesional de docentes (Ravela et al., 2007).

5.3 Aspectos técnicos

Aun cuando existe una comunicación adecuada entre equipos de trabajo de evaluadores y los operadores del sistema educativo, prevalece todavía en la región una falta de claridad sobre los propósitos y fines de las evaluaciones (Mosqueira y Robinson, 2018). Esto se atribuye, por un lado, a la falta de consenso sobre los propios fines y estándares de la educación y el currículum, que afecta la definición de qué se debe evaluar y para qué fines. Por otro lado, pocos actores comprenden que el uso de los resultados se desprende y define desde el momento del diseño técnico de la evaluación, no al revés; y que una evaluación no puede servir para todos los propósitos (Ravela, et al., 2007; Ravela, 2005):

Siempre que ambas finalidades [consecuencias “fuertes” y enfoque formativo] están presentes en un mismo acto de evaluación, las consecuencias fuertes predominan por sobre la finalidad formativa y la desvirtúan (Ravela, 2015, p. 103).

Este desconocimiento técnico sobre las evaluaciones, en ocasiones, lleva a que la urgencia de los tiempos políticos atropelle los tiempos requeridos para establecer propósitos claros y asegurar la calidad técnica de las evaluaciones (Ravela, et al., 2007).

Frecuentemente se contraponen visiones y argumentos opuestos respecto al objetivo de las evaluaciones que generan insatisfacción en los usuarios: resultados generados con fines de calidad se critican por su falta de enfoque formativo; o el rigor de la investigación no necesariamente atiende las necesidades de orientación pedagógica (Fortín Morales, 2013).

El comentario de una autoridad educativa en México ilustra este punt

Me gustaría tener más ayuda del INEE para hacer ese tipo de análisis coyunturales en materia de política. No la tenemos porque ellos se manejan en otro esquema, más académico, necesitan tiempo para analizar el problema, plantear y comprobar las hipótesis. Tenemos esquemas muy diferentes acerca del análisis de información y en el uso (...) (Silva et al., 2012, p. 69).

Relacionado con la tensión entre la lógica y tiempos de las unidades de evaluación y los de las áreas operativas del sistema educativo, algunos países como Perú, Chile y México han preferido apostar a levantamientos periódicos y regulares, con suficiente distancia en el tiempo, que les permita procesar y entregar resultados de manera más oportuna, sin sacrificar su calidad técnica (Ferrer, 2006; Silva et al., 2012).

Ante la expectativa de utilizar los resultados de las evaluaciones estandarizadas para informar políticas e intervenciones, se imponen las propias limitaciones de los levantamientos de este tipo. Los resultados de evaluaciones a gran escala, sobre todo cuando son censales, dificultan la recolección de factores del contexto escolar, social y del propio sistema, que se requieren para contextualizar adecuadamente los resultados en cada país o localidad y diseñar intervenciones acordes (Jiménez Moreno, 2016; Ferrer y Fiszbein, 2015). Los resultados de las evaluaciones, por sí mismos, pueden identificar un problema, pero no dan pistas sobre su posible solución. Sandoval-Hernández (2016), citado también por Jiménez Moreno (2016), argumenta que los resultados de PISA, por sí solos, pueden utilizarse para sustentar todo tipo de políticas, incluso contradictorias entre sí.

Esta limitante, además, genera interpretaciones simplistas y descontextualizadas que son ampliamente criticadas en la literatura. El informe del PREAL reporta que, para el año 2007, aún perduraban prácticas de difusión de resultados a nivel regional, del sector escolar o de escuelas, sin información sobre las condiciones y características del contexto, lo cual llevaba a interpretaciones aisladas o erróneas (Ravela et al., 2007). Jiménez Moreno (2016) retoma de Backhoff el ejemplo de México, donde la conversación mediática no ha considerado la inclusión en el sistema educativo de alumnos que anteriormente no asistían a la escuela al momento de interpretar las tendencias en los resultados de PISA que muestran un estancamiento aparente del sistema educativo.

Refiriéndose a prácticas existentes en 2006, Martínez Rizo comentaba:

Hoy las aplicaciones proliferan y comienzan a cansar a las escuelas; los resultados se debaten, a veces con escaso conocimiento de su complejidad. Para que las perspectivas favorables puedan concretarse será necesario que una visión objetiva de las posibilidades de la evaluación prevalezca sobre las visiones superficiales (2008, p. 26).

Rivas comparte esta postura al lamentar la proliferación de inexpertos educativos que opinan sobre la educación a partir de resultados de evaluaciones, y que han generado una “ola de desgaste sobre los maestros” (Rivas, 2015, p. 98).

El informe de 2015 del Banco Mundial aún menciona como reto pendiente este tema de la profundización en la conversación sobre resultados de evaluaciones (Ferrer y Fiszbein, 2015).

En respuesta a las inquietudes sobre la contextualización adecuada de los resultados, Ravela et al. (2007) proponen la realización de estudios de valor agregado, pero reconocen la falta de capacidad técnica en la región en este tema.

Refiriéndose de manera concreta al uso de resultados, Fortín Morales (2013) reconoce que hay poca capacidad técnica para utilizar los resultados en el aula, y no queda claro a quién corresponde desarrollar esa capacidad. Ravela et al. (2007) coinciden, especificando que la mayoría de los docentes desconoce lo que las pruebas miden y, por lo tanto, no saben cómo utilizar adecuadamente los resultados. Ferrer y Fiszbein (2015) añaden como una barrera el hecho de que los docentes y las familias no están preparados para leer los formatos en que se presentan los resultados de evaluaciones (barras, gráficos, escalas, tablas comparativas, textos discontinuos), a pesar de los esfuerzos de las instancias evaluadoras por simplificar la información y los reportes.

El informe del PREAL reporta una baja capacidad de los equipos técnicos de los ministerios para responder a los problemas que se identifican en las evaluaciones y, salvo en el caso de Brasil y Argentina, Rivas identificó una falta de atención a la formación técnica de funcionarios del sistema educativo (Ravela et al., 2007; Rivas, 2015). Como agravante del problema, la

alta rotación de personal en los ministerios de educación dificulta fortalecer el conocimiento y el manejo de las evaluaciones, así como la vinculación entre instancias evaluadoras y operativas (Silva et al., 2012).

Aunque el objeto de esta revisión no es abordar en detalle los aspectos técnicos de las pruebas, a continuación se presenta un esquema con la relación que establecen los autores del estudio del PREAL, entre el nivel de la calidad técnica de las pruebas, la difusión de los resultados y los usos y prácticas que derivan de ellos. El contenido de la tabla resume y explica, en gran medida, los fenómenos observados en la región en torno al uso de resultados.

Tabla 10. Relación entre calidad técnica y difusión de pruebas

DIFUSIÓN Y USO	CALIDAD TÉCNICA	
	BUENA	MALA
Adecuados a las características técnicas de la prueba	Aporte. Posibilidad de impacto en la mejora del rendimiento en las áreas evaluadas.	Efectos perversos, interpretaciones y decisiones erradas.
No adecuados a las características técnicas de la prueba	Efectos perversos, interpretaciones y decisiones erradas.	Efectos perversos, interpretaciones y decisiones erradas.
Escasos o nulos	Desperdicio de recursos y oportunidades.	Desperdicio de recursos.

Tomado de "Las evaluaciones educativas que América Latina necesita" (Ravelo et al., 2007, p. 10).

5.4 Comunicación y difusión

Finalmente, se debe mencionar el rol que juegan los mecanismos de reporte y difusión de los resultados sobre su uso.

El estudio del Banco Mundial de Ferrer y Fiszbein (2015) reconoce que ha habido un avance importante durante los últimos diez años en estrategias de comunicación de resultados para distintas audiencias, aunque queda mucho por hacer. En el mismo reporte se identifican las siguientes formas de presentar los resultados y las sugerencias para su interpretación:

- Informes completos con resultados nacionales y subnacionales.
- Informes de progreso longitudinal.
- Informes individuales para estudiantes y escuelas (en contextos de altas consecuencias).
- Bases de datos.

- Análisis secundarios de los datos.
- Recomendaciones pedagógicas a partir de ítems liberados.
- Motores de búsqueda en línea para encontrar resultados y ejemplos de contenidos evaluados.
- Presentaciones audiovisuales en distintos formatos.

De acuerdo con Fortín Morales, los retos de difusión deberían considerar que:

a) la información llegue a los usuarios; b) la información se represente de una manera que los usuarios la puedan comprender; c) la información proporcionada sea la necesaria para iniciar cambios a favor de la calidad educativa (2013, p. 17).

El mismo autor reclama una comunicación orientada explícitamente a niveles operativos del sistema educativo.

En este sentido, existe una conciencia generalizada de la necesidad de diversificar el tipo de productos para referirse a distintas audiencias y ya están en marcha varias iniciativas para atender el tema. Ayala y Sepúlveda (2018) resaltan los esfuerzos recientes que se han realizado en Chile para escribir reportes, breviaros de política e informes que ofrezcan sustento a tomadores de decisiones. Bracho (2018) reconoce que la oferta de información del INEE en México todavía es muy uniforme y poco diferenciada, mientras que otros estudios constatan que antes del 2013 existían en el país pocos canales de distribución de los reportes y resultados a nivel estatal (Silva et al., 2012).

Se percibe, en general, que los países con sistemas de resultados de consecuencias altas o “fuertes” dedican mayores esfuerzos a su difusión (Ferrer, 2006).

No solamente es necesario diversificar los formatos de difusión, sino atender a su calidad técnica y comunicativa para el público al que va dirigida. Varios países de la región han buscado innovar en el contenido de los reportes a escuelas para hacerlos más didácticos y sencillos; la inclusión de descripción de niveles de logro presenta un potencial de uso pedagógico cuyos resultados aún no se han documentado (Ferrer y Fiszbein, 2015). En

Honduras, se encontró que los reportes publicados a nivel municipal generaban mayor apropiación y respuesta que los publicados con resultados a nivel nacional (FEREMA, 2017).

Adicionalmente, en toda la región se han realizado esfuerzos por educar a las audiencias usuarias de resultados de evaluaciones (Ferrer y Fiszbein, 2015). En Chile, por ejemplo, se llevan a cabo seminarios y talleres con profesores, directores y actores de opinión pública para colocar temas de discusión en torno a los resultados de evaluaciones (Ayala y Sepúlveda, 2018). En Colombia, el ICFES entrega reportes a las Entidades Territoriales Certificadas (ETC) con los resultados de todas las escuelas bajo su adscripción con una guía de interpretación y uso de resultados, y se han realizado iniciativas de talleres para discutir los resultados en las escuelas (Rivas, 2015). Uruguay realiza talleres con profesionales de la educación para discutir los resultados y entrega guías de uso a las escuelas (Rivas, 2015).

No se puede omitir una referencia a las páginas web de los institutos o unidades de evaluación. La revisión de Ferrer y Fiszbein (2015) encuentra que en su mayoría no están bien diseñadas y presentan a menudo fallas técnicas. La estructura de las páginas no tiene un fin comunicativo claro y es difícil encontrar los reportes, datos o informes. La información se encuentra repetida y los nombres de las ligas no son intuitivos para audiencias no expertas. Resaltan como excepciones la página del INEVAL, de Ecuador, y la recientemente renovada página del INEE²⁷, que se dirigen de manera explícita a distintas audiencias (INEVAL, 2019; INEE, 2019).

6. Posturas frente a las evaluaciones estandarizadas

Este último apartado de la revisión busca recuperar las distintas voces que han surgido en la conversación en torno a los resultados de las evaluaciones en la región. Se presentan opiniones encontradas, críticas, defensas y algunas recomendaciones agrupadas en dos grandes categorías:

- Teórico-ideológicas: trabajos que abordan, desde posturas críticas, las

²⁷ Durante la elaboración de este trabajo se constató que frecuentemente los portales de internet se modifican con los cambios de administración, en ocasiones desechando esfuerzos previos que facilitaban el acceso a usuarios.

definiciones y los marcos conceptuales de las evaluaciones de gran escala en el marco de los sistemas de evaluación para medir la calidad educativa; así como cuestionamientos a las teorías de cambio que subyacen al uso de los resultados de las evaluaciones.

- Técnico-instrumentales: referencias a cuestionamientos y reflexiones derivadas de la implementación de las evaluaciones de gran escala y sus efectos sobre los sistemas educativos en la región.

6.1 Teórico-ideológicas

La utilización del término “calidad”, como referente para la evaluación, ha sido fuertemente criticada debido a la falta de consenso en torno a la definición de la calidad educativa.

No existe una postura única sobre lo que se entiende por “ideal educativo” y, por consiguiente, por “calidad”, pues esto depende de la corriente pedagógica e ideológica que adopta un sistema educativo en un determinado momento (Mosqueira y Robinson, 2018; Fortín Morales, 2013). Casassus (2007) denuncia que, ante este vacío, el término “calidad” en educación se ha reducido a un enfoque neoliberal dirigido a servir al mercado laboral, y de forma operativa, a aquello que se puede medir en pruebas estandarizadas. Esta visión es compartida por Jiménez Moreno (2016), quien critica la influencia de PISA en definir la calidad de la educación en función del dominio de habilidades y competencias para insertarse en el mercado laboral.

Prats (2011) señala que instrumentos como PISA se han constituido en evaluaciones del capital humano de cada país en el contexto de la competencia internacional y que esta asociación puede influir, negativamente, en las finalidades mismas de la educación.

En este sentido, las críticas más fuertes recaen sobre su influencia, no solamente en términos de definición de referentes educativos para los países, sino en marcar una agenda global más amplia para el desarrollo y, en concreto, para la educación (Casassus, 2007; Jiménez, 2017; Jiménez Moreno, 2016; Sánchez Cerón y Corte Cruz del Sagrario, 2013).

Más allá de la discusión en torno al término “calidad”, pero relacionado con la importancia de llegar a definiciones teóricas a nivel del sistema educativo, los autores del informe del PREAL (Ravela et al., 2007) argumentan

que la falta de claridad en metas y estándares de aprendizaje, así como en perfiles de egreso de estudiantes en distintos niveles, dificulta no solamente la propia evaluación, sino también la enseñanza. Se unen a la advertencia de los riesgos que conlleva equiparar las pruebas estandarizadas con medidas de la calidad de la educación o del sistema educativo.

Aunado a este último punto, Ravela subraya la importancia de ampliar la noción misma de “resultados”, ya que:

Los “resultados” de la evaluación educativa no son lo mismo que los resultados de la labor educativa (...) Los ‘resultados’ –o productos– de la evaluación deberían incluir una visión y una valoración del conjunto de la situación educativa, así como el análisis de los procesos que tienen lugar en el sistema educativo y que están en la base de sus logros e insuficiencias (Ravela, 2015, p. 84).

Como una alternativa, en esta línea comienzan a surgir medidas complementarias e índices que consideren variables adicionales (Bracho, 2018). Como se mencionó anteriormente, Brasil y Colombia están abriendo el camino con el IDEB y el ISCE, respectivamente (Ferrer y Fiszbein, 2015).

Otra crítica a este enfoque de calidad en las evaluaciones de gran escala es que obvia el resto de los componentes y recursos, las características, trayectorias y otras habilidades de los alumnos, así como los procesos y limitaciones del sistema educativo, y la propia responsabilidad del Estado frente a la tarea educativa (Mosqueira y Robinson, 2018; Jiménez Moreno, 2016; Ravela, 2015).

Algunas consecuencias son el deslinde del Estado frente a su rol educativo, la transferencia del peso de la responsabilidad del aprendizaje a los docentes y del control de las escuelas a las familias, además del intento de suplantar las funciones de acompañamiento y actualización que corresponden a supervisores u otras figuras con la promoción del enfoque “formativo” de las pruebas (Ravela, 2015; Jiménez, 2017).

Al respecto, el estudio del PREAL también advierte sobre el peligro de que, a través de sistemas de ranking o políticas de incentivos docentes con base en resultados, el Estado transfiera su responsabilidad educativa a las escuelas, docentes y familias. En este escenario, el Estado se limita a evaluar

e informar sin cumplir con su obligación de desarrollar capacidades para la mejora, dando prioridad a los contextos más desfavorecidos (Ravela, et al., 2007).

Tanto Casassus (2007) como Jiménez (2017) y Ravela (2015) denuncian la culpabilización de los docentes por los bajos resultados; y el último alude a los resultados de un estudio en Uruguay en el que los docentes sienten la exigencia de resolver los problemas sociales y de desarrollo económico a través de su función educadora mientras que no reciben las herramientas adecuadas ni se valora su profesión. Ejemplos de esta presión también se han reportado en Brasil, Chile y México (Sánchez Cerón y Corte Cruz del Sagrario, 2013).

6.2 Técnico-instrumentales

Desde su creación, los sistemas de evaluación generaron gran expectativa como recursos que permitirían diseñar políticas e intervenciones para la mejora basadas en evidencia. A casi 30 años del inicio de su implementación, es claro que las evaluaciones de aprendizajes, por sí mismas, no detonan procesos de mejora.

Surgen algunas críticas al modelo de evaluación de procesos arguyendo que éste es útil como modelo explicativo, pero no informa sobre cómo generar condiciones de mejora. Si bien, de manera general, se valora positivamente investigar sobre factores asociados al logro, algunos teóricos sostienen que no hay claridad sobre cómo interpretar la causalidad (Fortín Morales, 2013).

Casassus (2007) expresa su desencanto ante las expectativas generadas por las evaluaciones y los pocos resultados de mejora en la región, considerando la inversión que se hace en estas pruebas. Desde su postura, propone la supresión de las pruebas y que esos recursos se destinen directamente a mejorar el sistema.

En respuesta, otras posturas defienden la necesidad de contar con información sobre los logros de aprendizaje, reconociendo su función como un elemento más dentro de un sistema integral de evaluación y de los propios sistemas educativos (Mosqueira y Robinson, 2018; Martínez Rizo, 2008). A partir de la experiencia de Bolivia, donde no se aplican evaluaciones

estandarizadas desde el año 2000²⁸, Urquiola (2011) expresa las dificultades para valorar la calidad de la educación y el efecto de la inversión en el sector educativo sin datos sobre el logro educativo, y defiende la reinstalación de pruebas de gran escala. Sánchez (2018) advierte sobre los peligros de elaborar políticas sin evidencias, donde priman argumentaciones del gobierno con base en anécdotas o retórica, y donde las acciones dependen de los promotores de las políticas.

Buscando un punto de encuentro, el informe del PREAL expresa lo siguiente:

La evaluación es apenas uno de varios elementos de la política educativa que deben estar adecuadamente alineados: la formación docente inicial y en servicio, la carrera y condiciones para el trabajo docente, la gestión escolar, la supervisión, los diseños curriculares, los libros y materiales educativos, una inversión de recursos proporcional a las necesidades de las diversas poblaciones y una acción decidida por parte de los responsables educativos para resolver los problemas detectados, entre otros (Ravela et al., 2007, p. 5).

El texto plasma con claridad la complejidad que supone articular procesos de mejora, y otorga perspectiva sobre el rol viable y las limitantes propias de las evaluaciones estandarizadas.

A nivel de políticas públicas, si bien los resultados de estas pruebas resultan útiles para realizar un diagnóstico, Jiménez Moreno sostiene su insuficiencia para informar el diseño de políticas educativas:

(...) debido a que, para realizar esta tarea, se requiere la consideración de una gran cantidad de aspectos del contexto escolar y social que, por sus propias limitaciones, las pruebas no son capaces de mostrar (2016, p. 12).

Pensando en el uso formativo de las evaluaciones en las escuelas y aulas, también es importante reconocer sus limitantes:

²⁸ Desde creación en 1996, el Sistema de Medición de la Calidad (SIMECAL) aplicó pruebas en 1997 y en 2000, hasta su desmantelamiento en 2004 (Martínez Rizo, 2008). En 2017 el país retomó su participación en los estudios del LLECE con la aplicación del estudio preliminar denominado Pos-TERCE, coordinado por el LLECE y el Ministerio de Educación.

Los resultados de las pruebas estandarizadas no reflejan las complejas dinámicas educativas que operan al nivel de las escuelas y salones de clase, y en razón de ello, son insuficientes para orientar las prácticas de enseñanza (Martínez Rizo, 2016).

Es necesario superar la lógica tecnocéntrica del uso de resultados que no reconoce el proceso, las influencias y las tensiones de la elaboración de política. Se necesita considerar la complejidad del cambio educativo y definir el papel que la evaluación puede jugar en él: tiene la capacidad de mostrar áreas de mejora, pero no muestra el camino a seguir (Ravela, 2015).

Ravela sugiere que, más que esperar que el cambio provenga de la propia evaluación, ésta debe servir para confrontar las políticas y señalar dónde deben realizarse cambios:

Necesitamos una evaluación que no reemplace las políticas educativas, sino que las confronte (...) en el sentido de sugerir rumbos y promover políticas más apropiadas. La evaluación debe contribuir a restituir el lugar y la responsabilidad de la política educativa y no debe aceptar sustituirla (2015, p. 81).

Otro punto de desencuentro es la consideración del costo-beneficio de aplicar las evaluaciones:

Como se puede ver, desde hace más de una década se realizan evaluaciones a gran escala que implican la inversión de una cantidad significativa de recursos, pero lo que todavía no resulta claro —y sí ampliamente polémico— es qué tan útiles han resultado al sistema educativo nacional y a la sociedad mexicana (Jiménez Moreno, 2016, p. 6).

Posturas más radicales proponen la eliminación de las evaluaciones y la asignación de ese recurso directamente a acciones de mejora del sistema educativo, señalando el retroceso o estancamiento en los resultados como una prueba del fracaso de estos sistemas de evaluación²⁹ (Casassus, 2007). Otros grupos, como sucede en Guatemala, demandan un mayor aprovechamiento de la información por parte del Ministerio, proponiendo su uso

²⁹ Anteriormente en el documento se menciona la advertencia de Jiménez Moreno (2016) respecto a este tipo de interpretaciones, que no consideran otros factores que influyen en los resultados. El autor presenta el caso de México, en el que la incorporación de poblaciones anteriormente excluidas del sistema puede también explicar que no se hayan visto los avances nacionales esperados en aprendizajes.

en políticas de acreditación de alumnos o certificación de centros (Fortín Morales, 2013).

Al respecto, Ravela et al. (2007) informan sobre las ventajas, desventajas y retos anticipados de elegir sistemas de evaluación de certificación (cen-sales) o de diagnóstico (generalmente, muestrales). Uno de los temas que consideran es el costo de unas y otras, pero, en todo caso, defienden que las evaluaciones de calidad requieren de inversión por parte del Estado. Denuncian la existencia de gobiernos que demandan evaluaciones útiles, de alta calidad y de bajo costo. Más aún, subrayan la importancia de considerar y balancear la inversión de la evaluación y la de difusión de sus resultados y promoción de uso.

Por su parte, Bracho (2018) propone desarrollar políticas de usos de resultados que contemplen a distintos actores desde el momento del diseño de las evaluaciones, para aumentar los usos efectivos de los resultados y así, contribuir a elevar la calidad de la educación.

Desde una perspectiva teórica del tema, refiriéndose a la participación en evaluaciones internacionales, Addey y Sellar (2019) plantean ampliar la conversación de costo- beneficio para considerar las motivaciones de participación (o no participación) en las pruebas. La discusión sobre los costos o beneficios de las evaluaciones incluiría no solamente el uso directo de los resultados sino también las capacidades técnicas generadas, la pertenencia a la comunidad regional/internacional, la legitimidad política y la búsqueda del sustento racional de políticas. Como ejemplo, mencionan el caso de Bolivia, que justificó su no participación en PISA como una preferencia por establecer sus propios referentes educativos sin influencias externas ni comparativos con otros países que no comparten su contexto ni ideario de desarrollo. En este sentido, el costo de prescindir de las evaluaciones se contrapone al beneficio -desde la perspectiva del país- de autodeterminar su agenda educativa.

Finalmente, existe una serie de críticas respecto a los efectos no deseados que han producido los diferentes sistemas de evaluación, de acuerdo con las lógicas que los han regido.

De manera general, se acusa a las autoridades de buscar atajos simplifi-cadores que pretenden acelerar los procesos educativos, implementando

políticas que han resultado perjudiciales para el sistema y los aprendizajes (Rivas, 2015; Ravela, 2015).

En primer lugar, se critica que los sistemas de evaluación con consecuencias altas aumentan la inequidad y perjudican a escuelas en contextos más desfavorecidos. En el caso de Chile, los resultados de evaluaciones juegan un papel decisivo en la matrícula y recursos de la escuela, puesto que, en teoría, los padres de familia eligen las mejores escuelas para sus hijos. Con el paso del tiempo, se han instalado prácticas de segmentación en la que las escuelas subvencionadas o particulares eligen a los mejores estudiantes y las escuelas municipales se van rezagando (Sánchez Cerón y Corte Cruz del Sagrario, 2013). Esta dinámica de selección aumenta las brechas sociales, ya que las escuelas de escasos recursos se centran en subir puntajes a costa de los aprendizajes, mientras que en las escuelas particulares se pagan cursos de preparación (Casassus, 2007). El rechazo a estas prácticas ha sido categórico por algunos sectores de la sociedad y, más recientemente, por la misma ACE. Ayala y Sepúlveda (2018) reconocen que se han generado reacciones en contra de las evaluaciones que piden, inclusive, su eliminación. Un ejemplo de esto es el movimiento “Alto al SIMCE”, conformado por académicos, profesores y estudiantes (Sandoval-Hernández, 2016). Sin mencionar casos específicos, el informe del PREAL desaconseja prácticas de selección de alumnos para mejorar los resultados de las escuelas y advierte contra la adopción de expectativas de aprendizaje diferenciadas por escuelas de distintos estratos sociales (Ravela, et al., 2007).

En 1998, Carol Weiss ya identificaba, en un plano internacional, que los aspectos evaluados suelen convertirse en los aspectos prioritarios. En este sentido, hoy todavía existe la preocupación sobre el descuido de otros aspectos educativos como las humanidades, los valores y el arte en favor de las asignaturas de lenguaje y matemáticas, promovido por las evaluaciones estandarizadas (Sánchez Cerón y Corte Cruz del Sagrario, 2013). Varios autores coinciden en que el énfasis excesivo en los resultados de las evaluaciones estandarizadas empobrece la enseñanza al generar una visión de túnel en el que se desprecian los temas no evaluados y de desarrollo personal, aun cuando estos son altamente valorados por los propios docentes (Casassus, 2007; Rivas, 2015; Sánchez Cerón y Corte Cruz del Sagrario, 2013).

En el plano pedagógico, se acusa a las evaluaciones estandarizadas de homogeneizar prácticas ignorando el contexto de los alumnos, y de eliminar la flexibilidad de metodologías e innovación de los docentes (Sánchez Cerón y Corte Cruz del Sagrario, 2013). Un ejemplo es el enseñar a docentes a elaborar reactivos de evaluaciones estandarizadas para que los alumnos estén familiarizados con ellos (Silva et al., 2012).

Algunos autores sostienen que estos sistemas pervierten la motivación intrínseca de los docentes y asumen que necesitan de coerción externa para trabajar bien (Casassus, 2007; Sánchez Cerón y Corte Cruz del Sagrario, 2013). Por otro lado, la evaluación del programa piloto de incentivos de docentes en Colombia mostró una mejora en los resultados de alumnos cuyos docentes recibieron incentivos. Cabe señalar, sin embargo, que no se documentaron las prácticas realizadas en estas escuelas (Cabal, Milena, y Castro, 2018). Ravela (2015) argumenta que, si bien el sistema no debe funcionar exclusivamente a partir de exigencias externas, la naturaleza humana reclama cierto nivel de exigencia desde fuera. Al parecer, es necesario ahondar más en las dinámicas de motivación extrínsecas e intrínsecas que pueden resultar más efectivas para los docentes en la región.

Por otro lado, cuando se asocian incentivos o altas consecuencias a las escuelas o docentes a partir de los resultados de sus alumnos, se generan prácticas de simulación o de ensayo para el examen. En un análisis comparativo entre Brasil, México y Chile se describen prácticas de simulación o trampa, entrenar a los alumnos a contestar los exámenes³⁰ y pedir a los alumnos menos preparados que no tomen la prueba (Sánchez Cerón y Corte Cruz del Sagrario, 2013). En el caso de PISA, se distribuían los materiales previamente para que los docentes los trabajaran con los alumnos, bajo la consigna de subir los resultados (Silva et al., 2012).

Asociar este tipo de consecuencias a los resultados de las evaluaciones [incentivos a docentes por resultados de alumnos] reorienta los propios fines de la educación, pues los actores se ven estimulados a optar por caminos de formación que los lleven a la resolución eficiente de las evaluaciones, antes que a cubrir los objetivos de los planes y programas educativos (Jiménez Moreno, 2016, p. 16).

³⁰ En México, hasta el 2013 se llevaban a cabo semanas de Pre-EXCALE en el que los alumnos practicaban con pruebas similares a EXCALE.

Se han encontrado opiniones distintas sobre el interés de los alumnos frente a las pruebas, que al parecer depende de las consecuencias que puedan tener para ellos (Fortín Morales, 2013), aunque es un tema en el que ha habido poca exploración.

7. Conclusiones

A partir de la literatura revisada en este ejercicio, se puede concluir que, si bien existe disparidad en la madurez de los sistemas de evaluación entre países, hay una tradición de evaluación instalada en la región con avances importantes en la transparencia de los resultados. El reto pendiente parece ser el uso de la información publicada para influir en la mejora de la educación.

Existen marcos de referencia sólidos, cada vez más detallados, para identificar los planteamientos y teorías de cambio que se atribuyen a las evaluaciones en el marco de la gestión del sistema educativo.

Desde el punto de vista empírico, estos marcos y teorías se han puesto a prueba, evidenciando algunas de sus limitaciones a partir de las consecuencias no intencionadas que se han generado, sea por elementos de su diseño técnico, del diseño de los esquemas de consecuencias, de las capacidades para el uso de resultados, de la difusión de éstos, o de las coyunturas e intereses políticos.

Si bien, por un lado, hay una sistematización bastante concreta de las capacidades y limitaciones técnicas en los países para incluir factores asociados al logro dentro de modelos, índices o evaluaciones más complejas, que superen información descriptiva - así como de los cuerpos técnicos de áreas operativas y a nivel de docentes sobre la comprensión de los resultados de pruebas y cómo pueden usar la información de acuerdo con sus funciones y ámbitos de acción-; también hay menos información sobre las interpretaciones y usos precisos que distintos actores hacen de estos resultados más allá de los ejercicios de diagnóstico o seguimiento.

La conversación respecto a la pertinencia de estos sistemas incluye puntos de vista opuestos, con una fuerte carga crítica, sustentada en gran parte por las consecuencias no intencionadas de las evaluaciones. Sin embargo,

emergen voces de convergencia hacia una concepción más integral y medida de los sistemas de evaluación como parte articulada de los sistemas educativos, con un concepto de calidad educativa más amplio.

La concreción de esta visión integral de las evaluaciones de gran escala requiere de un conocimiento pormenorizado de las dinámicas de operación y la interrelación entre actores del sistema educativo en todos los niveles. Reducir la conversación a aspectos técnicos o comunicativos sería ignorar la naturaleza compleja y sistémica del proceso de mejora educativo.

A partir de los trabajos revisados, surgen los siguientes interrogantes que convendría abordar en estudios futuros:

A un nivel teórico que permita clarificar el rol de las evaluaciones y su articulación dentro de los sistemas educativos:

- ¿Qué factores median las expectativas de distintos actores frente a las evaluaciones de gran escala y bajo qué lógicas de cambio están operando?
- ¿Qué tipos de evaluaciones o estudios se requiere hacer sobre las dinámicas entre componentes y actores del sistema para superar una concepción de ellos como insumos estáticos del proceso educativo?
- ¿Qué esquemas de motivación interna e incentivos externos favorecerían la implementación sistémica de una cultura de evaluación para la mejora a nivel escolar y del uso formativo de los resultados?
- ¿Cómo se conjuga lo anterior con las políticas de rendición de cuentas?

A un nivel más operativo y técnico que permita desentrañar las dinámicas del sistema para incidir sobre ellas:

- ¿Cómo se interrelacionan a nivel nacional los factores políticos, técnicos e institucionales en relación con el uso observado de los resultados de evaluaciones?³¹
- ¿Qué rol juegan actualmente los resultados de las evaluaciones en los distintos momentos de diseño, implementación y seguimiento a políticas y programas para los actores involucrados?

³¹ Existe información sobre las capacidades técnicas de docentes, pero no ha habido suficiente investigación a nivel nacional sobre las posturas, actitudes, relaciones y capacidades técnicas de actores técnicos de mandos medios y altos.

- ¿Cómo impacta o explica la diferencia de visión y funciones esperadas de la evaluación por parte de los actores involucrados, en el diseño y el uso que se hace de ellas?
- ¿Existen factores, adicionales a los identificados, que influyan en el uso observado de los resultados de evaluaciones?
- ¿Qué información adicional se requiere utilizar con los resultados de evaluaciones para poder incidir en políticas, programas y en el proceso de enseñanza- aprendizaje?

Bibliografía

- Addey, C. y Sellar, S. (2019). Is it worth it? Rationales for (non)participation in international large-scale learning assessments. Working papers, UNESCO, Education Research and Foresight.
- Ayala, C. y Sepúlveda, M. (2018). Vanguardia en evaluación y mejora educativa. La agencia de la Calidad de la Educación en Chile. (R. Red, Entrevistador) Instituto Nacional para la Evaluación de la Educación. Recuperado de: <https://www.inee.edu.mx/index.php/publicaciones-micrositio/blog-revista-red/637-blog-revista-red/blog-revista-red-articulos/3417-vanguardia-en-evaluacion-y-mejora-educativa-chile>
- Bracho, T. (2018). Uso de la evaluación: asunto pendiente. (D. i. trabajo, Ed.) Instituto Nacional para la Evaluación de la Educación.
- Cabal, J., Milena, D. y Castro, F. (2018). Programa de Incentivos Docentes: Impactos y recomendaciones en el corto plazo. Departamento Nacional de Planeación, Bogotá. Recuperado de: https://colaboracion.dnp.gov.co/CDT/Sinergia/Documentos/Evaluacion_incentivos_docentes_Documento_vf.pdf
- Casassus, J. (2007). El precio de la evaluación estandarizada: la pérdida de la calidad y la segmentación social. (ANPAE, Ed.) Revista Brasileira de Política e Administração da Educação, 23(1), 71-79.
- CNSP (2019). Lineamientos generales del Sistema de Asesoría y Acompañamiento a las Escuelas. Secretaría de Educación Pública.
- CPEUM (2019). Constitución Política de los Estados Unidos Mexicanos. Diario Oficial de la Federación. Recuperado de: https://www.dof.gob.mx/nota_detalle.php?codigo=5560457&fecha=15/05/2019
- EDUY21 (sf). La evaluación de aprendizajes en una agenda de transformación educativa. Libro blanco. Montevideo: EDUY21 Iniciativa Ciudadana. Recuperado de: <http://www.eduy21.org/Documentos/1-Evaluacion-y-transparencia.pdf>
- FEREMA (2017). Educación: Una deuda pendiente. Informe de progreso educativo Honduras. Tegucigalpa: Fundación para la Educación Ricardo Ernesto Maduro Andreu.
- Ferrer, G. (2006). Educational Assessment Systems in Latin America: Current Practices and Future Challenges. Washington, D.C.: PREAL.
- Ferrer, G. y Fiszbein, A. (2015). ¿Qué es lo que ha sucedido con los sistemas de evaluación de aprendizajes en América Latina? Lecciones de la última década de experiencia. World Bank Group.
- Fortín Morales, Á. M. (2013). Evaluación Educativa Estandarizada en Guatemala: Un camino recorrido, un camino por recorrer. Guatemala: Ministerio de Educación de Guatemala.
- Fundación para la Educación Ricardo Ernesto Maduro Andreu (2017). Informe de

- progreso educativo Honduras 2017. FEREMA.
- Hanushek, E. y Woessmann, L. (2010). The high cost of low educational performance. The long-run economic impact of improving PISA outcomes. OECD.
- ICFES (2017). Informe nacional de resultados. Colombia en PISA 2015. Instituto Colombiano para la Evaluación de la Educación, Bogotá D.C. Recuperado el 3 de junio de 2019
- INEE (2018). Planea: nueva generación de pruebas. ¿Qué evalúa? Recuperado de: https://www.inee.edu.mx/wp-content/uploads/2018/12/Planea_3.pdf
- INEE (20 de mayo de 2019). Instituto Nacional para la Evaluación de la Educación. Recuperado de: www.inee.edu.mx
- INEVAL (2018). Educación en Ecuador. Resultados de PISA para el desarrollo. Ineval, Quito, Ecuador. Recuperado de: <http://www.evaluacion.gob.ec/evaluaciones/pisa-documentacion/>
- INEVAL (2019). Instituto Nacional de Evaluación Educativa. Recuperado de: <https://www.evaluacion.gob.ec/>
- Jiménez Moreno, J. A. (2016). El papel de la evaluación a gran escala como política de rendición de cuentas en el Sistema Educativo Mexicano. (Universidad Autónoma de Madrid, & Red Iberoamericana de Investigación sobre Cambio y, Edits.) Revista Iberoamericana de Evaluación Educativa, 9(1), 109-126.
- Márquez Jiménez, A. (2017). A 15 años de PISA: resultados y polémicas. (ISSUE-UNAM, Ed.) Perfiles Educativos, XXXIX (156), Editorial.
- Kautz, T., Heckman, J., Diris, R., ter Weel, B. y Borghans, L. (2014). Fostering and Measuring Skills: Improving Cognitive and Non-cognitive Skills to Promote Lifetime Success. Paris: OECD Publishing. doi:<https://doi.org/10.1787/5jxsr7vr78f7-en>
- Ley 20529 (2011). Ley del Sistema Nacional del Aseguramiento de la Calidad de la Educación Parvularia, Básica y Media y su Fiscalización. Santiago: Congreso Nacional de Chile. Recuperado de: <http://bcn.cl/1uv5c>
- LINEE (2013). Ley del Instituto Nacional para la Evaluación de la Educación. Diario Oficial de la Federación. Recuperado de: http://www.dof.gob.mx/nota_detalle.php?codigo=5313842&fecha=11/09/2013
- LOEI (2011). Ley Orgánica de Educación Intercultural. Quito: Registro Oficial del Órgano de Gobierno del Ecuador. Recuperado de: https://oig.cepal.org/sites/default/files/2011_leyeducacionintercultural_ecu.pdf
- Loza, S. E. (2018). Análisis a la evaluación educativa estandarizada en Colombia. Ensayo. Bogotá, Colombia.
- Martínez Rizo, F. (2008). La evaluación de aprendizajes en América Latina. Distrito Federal, México: INEE. Recuperado de: <https://www.inee.edu.mx/wp-content/uploads/2019/01/P1C140.pdf>
- Martínez-Rizo, F. (2016). Impacto de las pruebas de gran escala en contextos de débil tradición técnica: experiencia de México y el Grupo Iberoamericano de PISA.

- Revista Electrónica de Evaluación Educativa RELIEVE, 22(1), 1-12.
 Recuperado de: http://www.uv.es/RELIEVE/v22n1/RELIEVEv22n1_M0.pdf
- Mexicanos Primero (2012). Ahora es cuando. Metas 2012-2024. Ciudad de México: Mexicanos Primero Visión 2030, A.C. Recuperado de: http://www.mexicanosprimero.org/images/stories/mp_recursos/mp_publicaciones_de_mexicanos_primero/AhoraEsCuando2012-2024MetasWEB.pdf
- Mineducación (2015). Colombia, la mejor educada en el 2025. Líneas estratégicas de la política educativa del Ministerio de Educación Nacional. Colombia.
 Recuperado de: <https://www.mineduacion.gov.co/1759/w3-article-356137.html>
- Mosqueira, A. y Robinson, D. (2018). Análisis sobre la Evaluación de la Calidad Educativa en América Latina: Caso Colombia. (Universidad Autónoma de Madrid, & Red Iberoamericana sobre Cambio y Eficacia Escolar, Edits.) Revista Iberoamericana de Evaluación Educativa, 11(1), 43-55. doi:<https://doi.org/10.15366/riee2018.11.1.003>
- OECD (2019). Meeting expectations? Reflections on PISA for Development from participating countries. OECD. Recuperado de: https://www.oecd.org/pisa/pisa-for-development/Brief_29_Meeting_expectations.pdf
- Porter, C. (2010). What Shapes the Influence Evidence Has on Policy? The Role of Politics in Research Utilisation. University of Oxford, Department of International Development. Young Lives.
- Prats, J. (2011). Qué explica el éxito mediático del informe PISA. Escuela (3,8999).
 Recuperado de: http://www.ub.edu/histodidactica/images/documentos/pdf/exito_mediatico_informe_pisa.pdf
- Ravela, P. (2005). Para comprender las evaluaciones educativas. Fichas didácticas. Montevideo: PREAL. Recuperado de: <http://repositorio.minedu.gob.pe/bitstream/handle/123456789/4999/Fichas%20did%C3%A1cticas%20para%20comprender%20las%20evaluaciones%20educativas.pdf?sequence=1&isAllowed=y>
- Ravela, P. (2015). El uso de los resultados de las evaluaciones para mejorar los procesos de enseñanza. En M. Poggi, Mejorar los aprendizajes en la educación obligatoria. Políticas y Actores (págs. 69-112). IIPE UNESCO Buenos Aires.
- Ravela, P., Arregui, P., Valverde, G., Wolfe, R., Ferrer, G., Martínez-Rizo, F., . . . Wolff, L. (2007). Las evaluaciones educativas que América Latina necesita. Resultados de grupo de trabajo, Programa de Promoción de la Reforma Educativa en América Latina y el Caribe (PREAL), Grupo de trabajo sobre Estándares y Evaluación, S/D. Recuperado de: https://www.researchgate.net/publication/28216807_Las_Evaluaciones_Educativas_que_America_Latina_Necesita
- Rivas, A. (2015). América Latina después de PISA. Lecciones aprendidas de la educación en siete países (2000-2015) (1a. ed.). (F. CIPPEC, Ed.) Ciudad Autónoma de Buenos Aires, Argentina.
- Sánchez Cerón, M. y Corte Cruz del Sagrario, F. M. (2013). Las evaluaciones estandarizadas: sus efectos en tres países latinoamericanos. Revista Latinoamericana de

- Estudios Educativos, XLIII(1), 97-124.
- Sánchez, H. (2018). Autonomía de los organismos de evaluación ¿voluntad o derecho? (INAP, Ed.) Estado abierto. Revista sobre el Estado, la administración y las políticas públicas, 2(3), 15-45. Recuperado de: <https://tecnoadministracionpub.files.wordpress.com/2018/11/estado-abierto-vol-2-n-3.pdf>
- Sandoval-Hernández, A. (2016). Experiencias internacionales sobre el uso de los resultados de evaluaciones educativas. Mexico: INEE.
- Secretaría de Evaluación Educativa. (2017). Aprender 2017. Informe de resultados Primaria. Ministerio de Educación de Argentina.
- Silva, M., D'Angelo, N., Fernández, L., Martínez, O. y Rodríguez, A. (2012). El uso de los resultados de las evaluaciones del INEE por las autoridades educativas. Mexico City: INEE.
- UNESCO (2014). A brief history of LLECE. UNESCO Office in Santiago Newsletter (14º). Oficina de UNESCO en Santiago. Recuperado de: <http://www.unesco.org/new/en/santiago/press-room/newsletters/newsletter-laboratory-for-assessment-of-the-quality-of-education-llece/n14/a-brief-history-of-llece/>
- Urquiola, M. (2011). Calidad y cantidad educativa en Bolivia: 1996-2010. Informe preparado para el Ministerio de Relaciones Exteriores de Holanda. Minister van Buitenlandse Zaken, Policy and Operations Evaluation Department (IOB), El Haya. Recuperado de: <http://archieff.iob-evaluatie.nl/sites/iob-evaluatie.nl/files/351%20subrapport%20Urquiola%20Bolivia.pdf>
- Urquiola, M. (2011). Calidad y cantidad educativa en Bolivia: 1996-2010. Informe preparado para el Ministerio de Relaciones Exteriores de Holanda. Bolivia. Recuperado de: <http://www.iob-evaluatie.nl/sites/iob-evaluatie.nl/files/351subrapportUrquiolaBolivia.pdf>
- 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. Recuperado de: <https://pdfs.semanticscholar.org/976f/28dfe161d086b2605af6c3e364a7bf8bd991.pdf>



Organización
de las Naciones Unidas
para la Educación,
la Ciencia y la Cultura



Instituto Internacional de
Planeamiento de la Educación
IIPE-UNESCO Buenos Aires
Oficina para América Latina

Planificar **la educación**, construir **EL FUTURO**