

¿QUÉ HA SUCEDIDO CON LOS SISTEMAS DE EVALUACIÓN DE APRENDIZAJES EN AMÉRICA LATINA?

Lecciones de la última década de experiencia

Guillermo Ferrer y Ariel Fiszbein

Introducción

El desarrollo de sistemas de evaluación de aprendizajes en América Latina recibió un impulso significativo a partir de la década de los noventa como parte de los esfuerzos para modernizar los sistemas educativos—a menudo con un fuerte apoyo del Banco Mundial y otras organizaciones internacionales.

El Diálogo Interamericano, a través de la iniciativa de PREAL, apoyó y monitoreó el crecimiento y el desarrollo de esos sistemas. En un volumen publicado por PREAL, Guillermo Ferrer (2006) evaluó el estado de los sistemas de evaluación de aprendizajes en la región en los años 2004-2005. El informe reconoció el importante progreso que los países de América Latina habían logrado en el establecimiento de sistemas de evaluación, pero también mostró fuertes diferencias en el nivel de desarrollo de esos sistemas, dependiendo de las características de las unidades responsables de las funciones de evaluación y del contexto político en el que éstas funcionaban. El informe sugiere que las estructuras institucionales más estables fueron aquellas que trasladaron responsabilidades fuera de la estructura organizativa de los ministerios de educación a través de institutos con mayor autonomía administrativa y técnica. Según el informe, un área especialmente débil

en el funcionamiento de los sistemas de evaluación de aprendizajes fue la de difusión y utilización de los resultados. En particular, el informe argumenta que la baja participación de la sociedad civil en los procesos de evaluación no había promovido una demanda por mejorar los niveles de calidad de la educación.

La década desde la publicación de este informe ha sido una época de grandes cambios en la región. En primer lugar, en muchos países se han producido cambios significativos en las estructuras institucionales relacionadas con las funciones de evaluación de los aprendizajes. Por ejemplo, Ecuador estableció, como parte de una nueva constitución nacional, el Instituto Nacional de Evaluación Educativa (Ineval) en 2012 transfiriendo así responsabilidades que anteriormente estaban en manos del Ministerio de Educación (Mde). En México, el Instituto Nacional para la Evaluación de la Educación (INEE), que había sido creado por un decreto presidencial en 2002, adquirió un estatus jurídico más fuerte en 2013 a través de una nueva ley nacional de educación y una reforma constitucional.

En segundo lugar, han aparecido importantes novedades en la comunicación de los resultados, como por ejemplo, el Índice del Desempeño de la Educación de Brasil (IDEB) disponible al nivel de las escuelas, o los resultados de las pruebas SABER en Colombia, los cuales serán discutidos

Tabla de Contenidos

Introducción	1
Sistemas de evaluación de aprendizajes en América Latina: cambios y continuidades, 2005-2015	5
Marcos institucionales	5
Informes, difusión y uso de datos	7
Participación en pruebas internacionales	13
Lecciones aprendidas	14
Notas	17
Referencias	20
Anexo 1: Lista de personas entrevistadas	22
Anexo 2: Protocolo para las entrevistas (resumen de temas)	23
Anexo 3: Programas de evaluación 2009-2014	24
Anexo 4: Participación en pruebas internacionales	30

AGRADECIMIENTOS

Este trabajo ha sido posible gracias al apoyo del Banco Mundial y su Programa del Fondo Fiduciario READ (Russia Education Aid for Development) (disponible en: <http://www.worldbank.org/en/programs/read>). Los autores se beneficiaron de los valiosos comentarios de Marguerite Clarke y Gilbert Valverde, y el apoyo de Marianela D'Annunzio, María Oviedo, Ben Raderstorf, Laura Ruberto y Federico Sucre durante el proceso de elaboración del estudio.

más adelante en el presente informe. En tercer lugar, la participación de los países latinoamericanos en las pruebas internacionales también ha crecido a lo largo de la última década.

Quizás sorprendentemente, muy poco se ha escrito acerca de estos cambios. Además, no se sabe mucho acerca de los efectos que los cambios en los sistemas de evaluación pueden haber tenido en el sistema educativo y en la calidad de la educación en general.¹ Por ejemplo, mientras que el proyecto denominado ‘Metas Educativas 2021’—un conjunto de metas acordadas por los miembros de la Organización de Estados Iberoamericanos (OEI)—incluía una meta específica relacionada con el “fortalecimiento de los sistemas nacionales de evaluación”, el primer informe de seguimiento (OEI, 2012) sobre esos objetivos excluye explícitamente ese objetivo en particular de su evaluación, aparentemente debido a la falta de información. Cabe señalar, también, que de los 21 países para los cuales se dispone de un informe SABER sobre Sistemas de Evaluación de Estudiantes, ninguno está en América Latina.²

Creemos que suplir esta falta de información tiene gran un valor, tanto para los países de América Latina—que pueden aprender de la experiencia adquirida a lo largo de una década de intenso cambio—como para los demás

países en desarrollo, sobre todo teniendo en cuenta que América Latina fue la primera región en la que hubo una fuerte demanda conjunta para establecer sistemas de evaluación de aprendizajes. El creciente interés por medir los resultados de los aprendizajes en el contexto de las discusiones sobre los objetivos de desarrollo internacional después del 2015 (Naciones Unidas, 2014) y la consiguiente necesidad de fortalecer los esfuerzos nacionales para evaluar el aprendizaje hace que éste sea un ejercicio oportuno y, esperamos, útil.

Este trabajo es el resultado de una actualización (parcial) del informe de 2006; fue llevado a cabo durante un período de tres meses, con el propósito de ofrecer un análisis crítico de los progresos realizados y las lecciones aprendidas de la experiencia durante la última década. En lugar de examinar los progresos en todas dimensiones incluidas en el análisis original,³ siete en total, nos centraremos aquí sólo en tres dimensiones:

- (i) el marco institucional y organizativo de los sistemas de evaluación de aprendizajes
- (ii) la difusión y usos de la información (para/por las autoridades, actores escolares, familias, sociedad civil)
- (iii) la participación en las pruebas internacionales

TABLA 1: INDICADORES EDUCATIVOS BÁSICOS

	PIB PER CÁPITA (2013)	% DEL PIB DESTINADO A LA EDUCACIÓN (2013)	NÚMERO DE DOCENTES	NÚMERO DE ESTUDIANTES
Brasil	11.208	6,1*	2.515.648	46.512.407
Colombia	6.807	4,9	430.410	10.927.876
Guatemala	3.478	2,8	228.072	4.167.443
México	10.307	5,1*	1.932.061	32.077.094
Perú	6.662	3,3	493.284	9.282.045

*Datos de 2011

Fuentes: PIB per cápita y el % de PIB destinado a educación: Datos del Banco Mundial. A menos que se indique lo contrario, los datos son de 2013. Excepción: % del PIB de Brasil destinado a la educación se obtuvo de OCDE Education at a Glance Country Note. Número de estudiantes: Banco de datos de UIS. Incluye estudiantes en niveles 0 a 3 de CINE. Número de docentes: Perú: “Magnitudes” base de datos de ESCALE (Estadística de la Calidad Educativa) del Ministerio de Educación de Perú. Incluye docentes de EBR (Educación Básica Regular) en la escuela pre-primaria, primaria y secundaria en 2014. Guatemala: CIEN y Diálogo Interamericano (2015). El estado de las políticas públicas docentes, Guatemala. Incluye pre-primaria, primaria y profesores de secundaria de todo tipo de instituciones. México: SEP (2015). “Descarga serie histórica y pronósticos de la estadística del sistema educativo nacional”. Web. Secretaría de Educación Pública. Incluye los docentes en la escuela pre-primaria, primaria y secundaria en todo tipo de instituciones. Brasil y Colombia: “Maestros por nivel de educación”. [Conjunto de datos]. UIS Instituto de Estadística de la UNESCO.

Dado el escaso tiempo disponible para este ejercicio, centramos nuestra atención en la experiencia en cinco países: Brasil, Colombia, Guatemala, México y Perú (ver **Tabla 1**). Nuestra selección de casos de países trató de capturar la experiencia de algunos de los países que han realizado importantes inversiones⁴ en los sistemas

En la última década los sistemas estandarizados de evaluación de aprendizajes a gran escala en América Latina han ganado visibilidad, fuerza y un papel prominente en la discusión general sobre la calidad educativa.

de evaluación de aprendizajes durante este período. Al mismo tiempo, hemos querido capturar la diversidad de condiciones en estos países. En particular, hemos tratado de cubrir tanto países en los que las funciones de evaluación son responsabilidad de instituciones que funcionan de manera independiente, como aquellos en los que la función se lleva a cabo por unidades que operan dentro de la estructura formal del Ministerio de Educación. Los países cubiertos también representan una cierta diversidad regional (incluyendo países en Norteamérica, Centroamérica y Sudamérica). El caso de Guatemala es útil como un ejemplo de país que históricamente ha contado con un fuerte apoyo técnico y financiero por parte de organismos internacionales.

Para estos países se realizó un análisis en profundidad y una revisión exhaustiva de materiales oficiales (disponibles a través de sitios web oficiales)⁵ y un conjunto de entrevistas semi-estructuradas con doce expertos⁶ en la operación de los sistemas de evaluación de aprendizajes en estos países. Se hizo un esfuerzo para entrevistar tanto a personas que han jugado un papel clave en la gestión de

estos sistemas (pero que ya que no están directamente involucrados con esos sistemas de manera oficial) como así también a personas que representan a la sociedad civil independiente, la academia y las voces técnicas en el campo de la evaluación educativa.

Más allá de los cinco países en los que se puso el foco de atención, el presente informe incluye el análisis de información sobre la participación en pruebas internacionales para toda la región. Este análisis también se ve beneficiado por nuestro conocimiento de lo que sucede en otros países de la región para los cuales, sin embargo, no se llevó a cabo una revisión de la documentación oficial ni se realizaron entrevistas.

Este ha sido un proyecto a corto plazo, y estamos conscientes de las limitaciones de nuestro análisis y de la necesidad de recopilar más información para probar aún más nuestras hipótesis. Probablemente la observación más coherente que se puede hacer en este momento, a modo de introducción, es que en la última década los sistemas estandarizados de evaluación de aprendizajes a gran escala en América Latina han ganado visibilidad, fuerza y un papel prominente en la discusión general sobre la calidad educativa. En general, la tendencia en América Latina es, sin lugar a dudas, mantener el apoyo a la labor de las unidades nacionales de evaluación⁷ y fortalecer sus componentes técnicos e institucionales.

Sin embargo, este progreso no ha sido automático. Para lograr este fortalecimiento, los países han luchado por mantener con firmeza tanto la forma como el contenido de lo que ya existía. Y en algunos casos, como se describe a continuación, hicieron cambios y ajustes para asegurar la estabilidad y continuidad de las acciones en tiempos de adversidad. Esta observación se aplica principalmente, y más directamente, al primero de los tres aspectos principales de evaluación abordados en este informe, es decir, el marco institucional para las evaluaciones. No obstante, también tiene implicaciones directas para el uso de los datos y de los resultados, así como para la participación de los países en las evaluaciones internacionales.

Sistemas de evaluación de aprendizajes en América Latina: cambios y continuidades, 2005-2015

En este trabajo nos propusimos evaluar los cambios y continuidades en tres dimensiones clave de los sistemas de evaluación de aprendizajes en América Latina tomando como referencia las conclusiones del informe de 2006. La primera dimensión que consideramos es el marco institucional para la función de evaluación, y sus efectos sobre la credibilidad y la legitimidad de las actividades de evaluación, así como la sostenibilidad técnica y operativa de dichas actividades. Esta dimensión de nuestro análisis se corresponde con lo que el marco de la Evaluación de Estudiantes SABER llama 'Contexto Facilitador'.⁸

La elección entre un modelo que se basa en los organismos autónomos (o semi-autónomos) y uno que utiliza unidades especializadas dentro del Ministerio de Educación para dirigir y administrar las evaluaciones de educación es el aspecto más notable en el que nos enfocamos. La segunda dimensión que consideramos es la presentación de informes, difusión y uso de los datos. La influencia que tengan las evaluaciones de aprendizajes en la calidad de la educación depende fundamentalmente de las formas en que se presentan los resultados, la información se comparte y los hallazgos son utilizados por las escuelas, maestros, administradores, padres de familia y la sociedad civil en general—y nosotros tratamos de comprender los cambios que han ocurrido en esos aspectos. Esta dimensión de nuestro análisis corresponde a lo que el marco de la Evaluación de Estudiantes SABER llama 'Calidad de la Evaluación'. Por último, consideramos la participación en pruebas internacionales, no sólo como un indicador del compromiso de un país con la evaluación sino también como un factor que influye la capacidad de evaluación que ese país tiene.

MARCOS INSTITUCIONALES

En los cinco casos nacionales analizados, el liderazgo y la responsabilidad administrativa de los programas de evaluación estandarizada recae en dos grandes tipos de instituciones: los institutos nacionales y unidades de evaluación dentro de los Ministerios de Educación (MdE, o de la Secretaría de Educación Pública, SEP, en el caso de México). El caso de México actualmente presenta una

situación que se puede caracterizar como un formato mixto, que consiste en un instituto nacional—el Instituto Nacional para la Evaluación de la Educación (INEE)—y la unidad de evaluación dentro de la SEP. Los otros dos institutos nacionales son el Instituto Colombiano para La Evaluación de la Educación (ICFES) en Colombia y el Instituto Nacional de Estudios e Pesquisas (INEP) en Brasil. Las unidades dentro de los MdE son la Unidad de Medición de la Calidad Educativa (UMC) en el Perú y la Dirección General de Evaluación e Investigación Educativa (DIGEDUCA) en Guatemala.⁹ Todos los institutos y unidades de los MdE ya existían en 2005, con la excepción de DIGEDUCA, que ya estaba operando para ese entonces, pero fue constituida oficialmente por ley en 2007.

Una serie de cambios se han producido en los institutos nacionales, sobre todo en términos de nuevas disposiciones legales que le otorgan a los institutos mayor autonomía administrativa. El ICFES, por ejemplo, ahora tiene más responsabilidades para ofrecer servicios de evaluación, recaudar ingresos, y como resultado de las reformas introducidas en 2009, retener excedentes para ser reinvertidos en mejoras técnicas y desarrollo de programas.

El INEE de México sufrió un cambio institucional importante cuando la organización, originalmente creada por un decreto presidencial en 2002, llegó a ser ratificado por la ley a través de una reforma constitucional aprobada en 2013. Este cambio le dio al INEE no sólo una mayor autonomía, sino también más responsabilidades, incluyendo el diseño y el liderazgo de un Sistema Nacional de Evaluación, que involucra a los programas de evaluación administrados tanto por el Instituto como por la SEP¹⁰. Cabe señalar que, a pesar de que se suponía que las operaciones del INEE debían reemplazar las evaluaciones de la SEP a partir del 2002, lo que realmente ocurrió fue que la SEP se fortaleció, y se hizo también mucho más visible, con la implementación de una evaluación censal que captó rápidamente la atención nacional, mientras que el papel y la influencia del INEE en la conducción de las funciones de evaluación educativas disminuyó. Debido a una serie de razones técnicas y políticas, el INEE ha recuperado tanto poder como autoridad legal para coordinar un Sistema Nacional de Evaluación, que incluye todas las actividades de evaluación, tanto dentro del Instituto como también las que realiza la SEP. De qué manera evolucionará este arreglo durante los próximos años, en particular, el equilibrio entre autonomía y rendición de cuentas, será sin duda un tema interesante para futuras investigaciones.

El INEP en Brasil ha sido una organización autárquica¹¹ durante muchos años y, mientras que los intentos de establecerlo como un organismo totalmente independiente no tuvieron éxito debido a un vacío legal en relación con los organismos públicos, el Instituto todavía tiene amplio poder discrecional para su autogobierno y administración.

En los tres casos, sin embargo, el Director de estos institutos es seleccionado por el Ministro de Educación y nombrado por el Presidente de la Nación (Colombia y Brasil) o por el Senado (México). En resumen, todos los Institutos han continuado funcionando durante la última década, y su autonomía y responsabilidades se han ampliado y fortalecido, ya sea por la práctica misma o a través de nuevas disposiciones legales.

A pesar de que la autonomía de los institutos no se cuestiona, y rara vez es políticamente desafiada, hay una demanda manifiesta y permanente de que su trabajo de evaluación esté cada vez más estrechamente alineado con las políticas educativas del gobierno.

Las unidades como la UMC y DIGEDUCA han continuado operando dentro de la estructura organizativa de los ministerios. Aunque todavía hay cierto debate sobre si deben ser convertidos en organismos autónomos fuera del MdE; la capacidad técnica y la transparencia que han demostrado a lo largo de los años han hecho que la cuestión del cambio institucional sea un tema de menor prioridad. Tanto en Guatemala como en Perú las unidades de evaluación operan casi en su totalidad con fondos del

presupuesto nacional y no de fuentes internacionales, a diferencia de lo que había ocurrido en el pasado. Como resultado, estas unidades se han independizado cada vez más de la ayuda y de los créditos internacionales. Esta independencia creciente de las agencias externas denota la presencia de la voluntad política nacional—por un conjunto de razones posibles—para mantener estas unidades operando plenamente, incluso cuando los resultados de evaluación puedan ser bajos y representen “una mala noticia” para las autoridades políticas y los responsables políticos. Esta no era una predicción obvia diez años atrás.

A pesar de que la autonomía de los institutos no se cuestiona, y rara vez es políticamente desafiada, hay una demanda manifiesta y permanente de que su trabajo de evaluación esté cada vez más estrechamente alineado con las políticas educativas del gobierno, particularmente en lo relacionado con la implementación del currículo y el desarrollo profesional de los docentes. Sin embargo, hay que reconocer que, a pesar de tener sus propias unidades de evaluación, los MdE también a menudo tienen dificultades para gestionar un flujo interno adecuado de los datos para mejorar la toma de decisiones sobre la base de los resultados de aprendizaje de los estudiantes. Esto se explicará con más detalle en la siguiente sección, pero por ahora basta con decir que ningún tipo de arreglo institucional puede garantizar por sí mismo, un uso eficiente de los datos para la elaboración de políticas informadas y para una mejora educativa.

Un aspecto en el que parece haber una diferencia importante entre los institutos y las unidades es la posibilidad de diseñar y llevar a cabo con eficacia planes de evaluación a mediano y largo plazo. Los Institutos tienden a tomar decisiones a largo plazo sobre los programas de evaluación (incluyendo la población a evaluar, la cobertura de contenido, los ciclos de administración de los exámenes, etc.) y se ciñen a ellos durante todo el tiempo que se los considere útiles y pertinentes. A las unidades de los MdE, por otro lado, les resulta mucho más difícil mantener planes de evaluación a largo plazo debido a frecuentes—y en algunos casos erráticas y arbitrarias—demandas de las autoridades políticas, en particular cuando se producen cambios de gobierno. Los Institutos también son propensos a recibir este tipo de demandas, pero definitivamente tienen mucho más poder y autoridad legal para discutir la conveniencia de los cambios y, en última instancia, decidir si serán o no implementados. En el caso de México, bajo su nuevo marco legal y constitucional, el INEE puede ahora no

sólo decidir libremente sobre sus propios programas de evaluación, sino también dictar lo que la SEP puede o no hacer en términos de evaluación educativa.¹²

Hay por lo menos otros dos factores políticos que podrían –en principio– distinguir a los institutos de las unidades dentro de los MdE:

- a. Uno de ellos se refiere a la “sensibilidad” de estas organizaciones a la interferencia política en el manejo de los datos públicos, es decir, los resultados de la evaluación. En la práctica, sin embargo, durante la última década no ha habido ninguna diferencia en este aspecto entre los países con diferentes marcos institucionales. Ya en 2005 los resultados siempre se hicieron públicos, y así sigue siendo. Los últimos dos casos notorios de censura de resultados ocurrieron en torno al año 2000, cuando México y Perú prohibieron la publicación de datos internacionales o nacionales de evaluación. A partir de 2005 y hasta el presente, no ha habido informes sobre intentos de censurar la difusión de datos en cualquiera de estos países y por lo tanto los factores de autonomía o de subordinación al gobierno no se han puesto a prueba en este sentido.
- b. Los sindicatos docentes, el otro actor político importante y detractor histórico de las evaluaciones estandarizadas, han desempeñado un papel esperado, esto es, se han opuesto y criticado pero esto sucede independientemente del marco institucional de la función de evaluación. Cuando han actuado o hecho pública su posición en contra de las evaluaciones, o cuando se han puesto de acuerdo para evitar la confrontación en esta materia, a cambio de un aumento en las compensaciones, lo han hecho en conversaciones directas con las autoridades políticas superiores. Por lo que se desprende de las entrevistas realizadas para este informe, son los Ministros de Educación y los Presidentes quienes toman la iniciativa en las negociaciones; los institutos o las unidades de evaluación de los MdE no son los objetivos o los sujetos de las disputas políticas en relación con la evaluación educativa.¹³

Otro aspecto organizacional que posiblemente podría implicar una distinción entre los institutos y las unidades de MdE, es su capacidad diferencial para crear y desarrollar experiencia técnica sólida y estable. El

tamaño y la autonomía administrativa de los institutos podrían llevarnos a pensar que tienen más oportunidades y son más capaces de reclutar y mantener un número permanente y creciente de expertos. La verdad es que, mientras que los institutos tienen más autonomía administrativa y estabilidad presupuestaria para contratar nuevo personal, el grupo de profesionales aptos para trabajar en puestos altamente especializados es bastante reducido en todos estos países. Por lo tanto, los institutos y las unidades de los MdE se enfrentan a un desafío permanente para seleccionar los candidatos adecuados, entrenarlos, y luego retenerlos dentro de la organización. Los profesionales idóneos generalmente provienen de los campos de la psicología, la sociología, la estadística y la economía (rara vez son especialistas en educación con formación pedagógica fuerte). Muchos de ellos, después de obtener el conocimiento y la experiencia en medición educativa, pasan a otras áreas de gobierno, a las empresas privadas o realizan estudios de posgrado en el extranjero. Además, las escuelas o facultades de educación en las universidades de la región no cuentan con suficientes programas (de alta calidad) en estadística y medición educativa. Tal vez una excepción a destacar es el INEP de Brasil, que a pesar de las dificultades en la contratación de personal nuevo y bien entrenado, al menos ofrece estabilidad en el empleo bajo las leyes de la administración pública local, un incentivo apreciado por muchos candidatos y empleados nuevos.

INFORMES, DIFUSIÓN Y USO DE DATOS

Los marcos institucionales, como se describen en la sección anterior, son importantes para el análisis en la medida en que éstos pueden afectar la producción y el uso de los datos, así como también la forma en que llegan a las audiencias y la forma en que se puede acceder a ellos públicamente. En esta sección se presenta una descripción de las diferentes formas en que los datos recogidos a través de la evaluación de aprendizajes se reportan, es decir, los modelos y formatos utilizados para la presentación de los datos, la forma en que se canalizan a diferentes audiencias, y cómo se utilizan o se espera que se utilicen para mejorar la calidad de la educación.

Todos los países que consideramos en este trabajo tienen una combinación de evaluaciones de altas y bajas consecuencias (“high-stakes” y “low-stakes”). Las evaluaciones de altas consecuencias, destinados por definición a tener consecuencias directas para las personas o instituciones (estudiantes, profesores, escuelas), determinan en gran medida la regularidad con la

cual se llevan a cabo, y también qué tan rigurosa debe ser la entrega de los resultados. Claramente, las evaluaciones (o más bien los exámenes) como SABER 11 en Colombia y ENEM en Brasil, los cuales tienen consecuencias para el acceso de los estudiantes a la educación superior, siempre se administran de acuerdo con un calendario riguroso, y sus resultados son entregados a escuelas y estudiantes en fechas previamente anunciadas en el calendario académico. Hay otras evaluaciones que, aunque cada vez más utilizadas en la práctica para proporcionar incentivos económicos a las escuelas con base en el progreso educativo, no requieren de un calendario de administración tan estricto como las anteriormente nombradas. Este sería el caso de las evaluaciones censales de educación primaria de Perú, la brasileña Prova Brasil o las pruebas SABER 3-5-7-9 de Colombia.¹⁴ Los últimos casos no estaban destinados originalmente a ser evaluaciones de altas consecuencias, pero se han convertido progresivamente en evaluaciones de gran visibilidad que ponen el foco en las escuelas, con el potencial para estimular o fomentar la responsabilidad social y política.

Por último, están las evaluaciones que no tienen consecuencias directas para las escuelas o para los

individuos, y que en su mayoría son muestrales o voluntarias, tales como Provinha Brasil, las evaluaciones de educación primaria de Guatemala, o las evaluaciones EXCALE de México desarrolladas por el INEE.¹⁵ En estos casos, sin embargo, la administración de pruebas también es bastante regular, de hecho más de lo que podría haberse esperado hace una década. No obstante, la regularidad de estas pruebas depende del calendario de las unidades de evaluación, de las posibilidades técnicas, de reglamentaciones legales o incluso de la voluntad política.

El formato en que se presentan los resultados sigue un patrón común independientemente de la naturaleza de las evaluaciones.¹⁶ Hace diez años, la mayoría de los países de América Latina ya habían transitado de la Teoría Clásica de los Tests (prevista para comparar, seleccionar, o clasificar a los individuos) hacia la Teoría de Respuesta al Ítem, que permite la comunicación de los resultados de lo que los estudiantes saben y pueden hacer, de una manera mucho más descriptiva y permite determinar qué tan básico o avanzado es su conocimiento. En otras palabras, independientemente de si las evaluaciones de hoy tienen el propósito de comparar o clasificar a los estudiantes en

GRÁFICO 1: EVOLUCIÓN DE IDEB PARA UNA ESCUELA SELECCIONADA AL AZAR

Fuente: Índice de Desenvolvimento da Educação Básica (IDEB) - Índice de Desarrollo de la Educación Básica, Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. (<http://www.qedu.org.br/escola/1186-escola-absolon-moreira/ideb>)

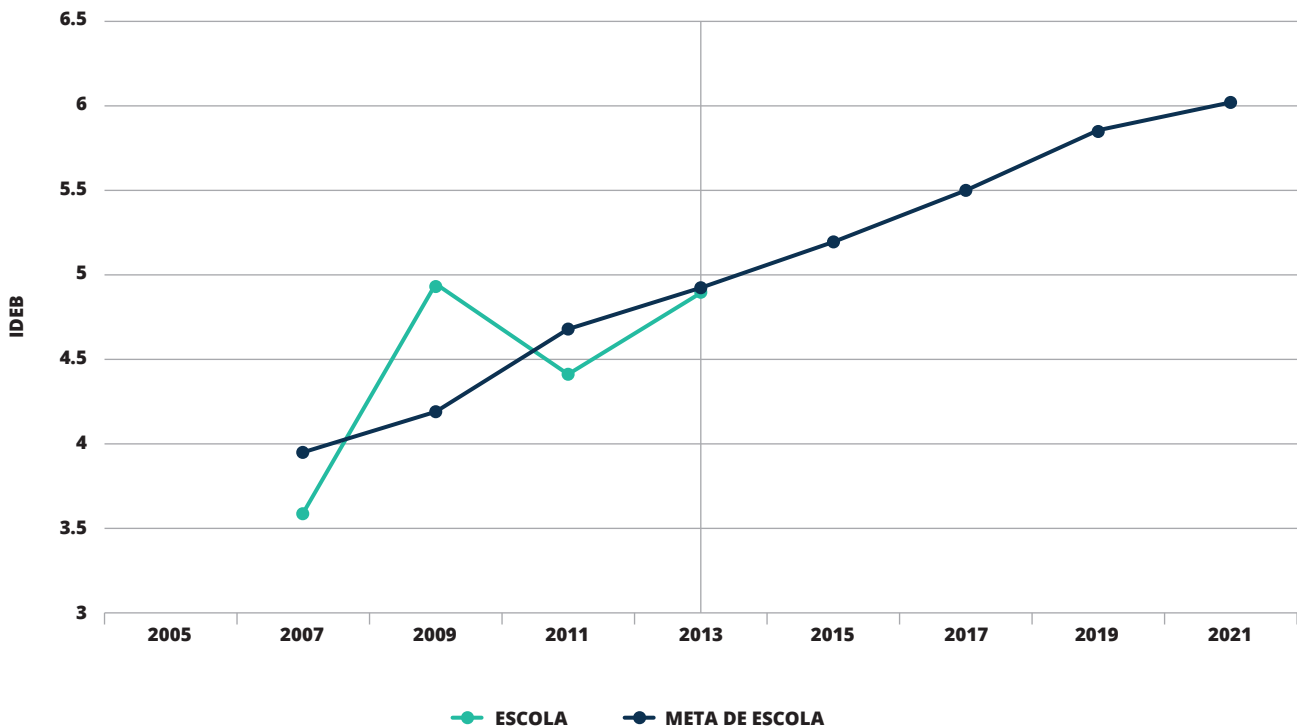


GRÁFICO 2: RESULTADOS Y METAS DEL IDEB

Fuente: Índice de Desenvolvimento da Educação Básica (IDEB), Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. (<http://ideb.inep.gov.br/resultado/>)

IDEB: 2005, 2007, 2009, 2011, 2013, Projeções para o Brasil

ANOS INICIAIS DO ENSINO FUNDAMENTAL										
	IDEB OBSERVADO					METAS				
	2005	2007	2009	2011	2013	2007	2009	2011	2013	2021
Total	3.8	4.2	4.6	5.0	5.2	3.9	4.2	4.6	4.9	6.0
DEPENDÊNCIA ADMINISTRATIVE										
Estadual	3.9	4.3	4.9	5.1	5.4	4.0	4.3	4.7	5.0	6.1
Municipal	3.4	4.0	4.4	4.7	4.9	3.5	3.8	4.2	4.5	5.7
Privada	5.9	6.0	6.4	6.5	6.7	6.0	6.3	6.6	6.8	7.5
Pública	3.6	4.0	4.4	4.7	4.9	3.6	4.0	4.4	4.7	5.8

ANOS FINAIS DO ENSINO FUNDAMENTAL										
	IDEB OBSERVADO					METAS				
	2005	2007	2009	2011	2013	2007	2009	2011	2013	2021
Total	3.5	3.8	4.0	4.1	4.2	3.5	3.7	3.9	4.4	5.5
DEPENDÊNCIA ADMINISTRATIVE										
Estadual	3.3	3.6	3.8	3.9	4.0	3.3	3.5	3.8	4.2	5.3
Municipal	3.1	3.4	3.6	3.8	3.8	3.1	3.3	3.5	3.9	5.1
Privada	5.8	5.8	5.9	6.0	5.9	5.8	6.0	6.2	6.5	7.3
Pública	3.2	3.5	3.7	3.9	4.0	3.3	3.4	3.7	4.1	5.2

ENSINO MÉDIO										
	IDEB OBSERVADO					METAS				
	2005	2007	2009	2011	2013	2007	2009	2011	2013	2021
Total	3.4	3.5	3.6	3.7	3.7	3.4	3.5	3.7	3.9	5.2
DEPENDÊNCIA ADMINISTRATIVE										
Estadual	3.0	3.2	3.4	3.4	3.4	3.1	3.2	3.3	3.6	4.9
Privada	5.6	5.6	5.6	5.7	5.4	5.6	5.7	5.8	6.0	7.0
Pública	3.1	3.2	3.4	3.4	3.4	3.1	3.2	3.4	3.6	4.9

un escenario de evaluación de alto impacto, o si están destinadas a proporcionar información sobre la calidad de su aprendizaje, siempre presentan escalas de logro, divididas en niveles de desempeño, mostrando no sólo el porcentaje de estudiantes en cada segmento, sino también una descripción sustancial de lo que los estudiantes saben y pueden hacer en cada nivel de la escala. Sin duda, esta metodología y formato de informe tiene mucho más potencial para las intervenciones pedagógicas que un informe sobre el porcentaje medio de preguntas respondidas correctamente. Vale la pena señalar que esta metodología más moderna y mucho más sofisticada se ha continuado usando en los últimos años, y esta tendencia está en sintonía con los últimos avances metodológicos en evaluación en todo el mundo.

Al mismo tiempo, las pruebas de altas consecuencias suelen ser censales. Esto tiene implicaciones para el análisis de la calidad de la educación, sobre todo en términos de cobertura curricular, es decir, cuánto pueden capturar las pruebas de lo que se enseña, o de lo que se supone que se enseña. Como es bien sabido, las evaluaciones censales cubren grandes poblaciones de estudiantes, y por lo tanto no pueden profundizar en la cobertura de los contenidos curriculares. Es por eso que se las caracteriza en la jerga de la evaluación, como evaluaciones “delgadas”.

Esto tiene implicaciones negativas por dos razones, ambas expresadas enfáticamente por los expertos en la materia. En primer lugar, como se ha indicado anteriormente, los resultados no permiten un análisis profundo del llamado currículo implementado. Los datos en este sentido son demasiado escasos. En segundo lugar, las evaluaciones censales, al ser tan “delgadas”, no permiten hacer un análisis riguroso y útil de los factores asociados al aprendizaje,¹⁷ como lo son las características de la escuela, el capital cultural de los estudiantes y de las familias o las metodologías de enseñanza.

Sin embargo, la última década muestra una serie de ejemplos de cómo las evaluaciones censales se pueden utilizar para el desarrollo de índices educativos de calidad, tales como el Índice de Desenvolvimento da Educação Básica de Brasil (IDEB) o el nuevo Índice Sintético de Calidad Educativa (ISCE) de Colombia. Estos índices representan uno de los últimos y más importantes avances en términos de uso de datos en la región (ver **Gráficos 1 y 2**).

En ambos casos, los índices se han desarrollado con dos propósitos principales. En primer lugar, para lograr una

medida de la calidad de la educación que tenga en cuenta al menos tres variables diferentes, no sólo los logros de aprendizaje. El IDEB de Brasil combina los resultados de aprendizaje con la tasa de aprobación anual de los estudiantes y las tasas de abandono, mientras que mide simultáneamente el progreso anual del índice en todas las escuelas de educación primaria del sistema.¹⁸ El ISCE de Colombia considera el rendimiento de los estudiantes en conjunto con la aprobación de año y la tasa de abandono, así como también el clima escolar. Este índice también mide el progreso anual de las distintas escuelas. En ambos casos, se proporcionan datos sobre los logros de aprendizaje, respectivamente, por las evaluaciones nacionales Prova Brasil y SABER. Los índices también permiten que los responsables políticos tengan una medida de progreso cuando establecen metas de logro a mediano y largo plazo. Por ejemplo, Brasil ha hecho pública su intención de aumentar los resultados anuales, en términos de aumento de los valores del índice, hasta el año 2021. Esto es claramente un signo de responsabilidad política, sin precedentes en América Latina hasta los inicios de esta década.

El segundo propósito para el cual se han desarrollado los índices es el de establecer programas de incentivos para aquellas escuelas que muestran un progreso en los valores del índice. Aún está por verse si estos incentivos tienen o tendrán un impacto positivo sobre los aprendizajes de los alumnos. Hasta el momento, no hay suficiente evidencia científica para demostrar qué tan adecuados o efectivos son estos incentivos. Sin embargo, el hecho de que se estén desarrollando mediciones más complejas es un signo positivo que muestra que el uso de los datos de evaluación está teniendo un rol más importante en estos países.

Para sostener esta conclusión, aunque sólo sea por contraste, es necesario tener en cuenta que los programas de incentivos que se basan exclusivamente en los datos sobre logros de aprendizajes son mucho más propensos a fomentar la corrupción y la manipulación de datos, como ha sido el caso de la evaluación de Enlace administrada por la SEP en México. En los últimos cinco años, los informes de fraude y corrupción han rodeado los exámenes Enlace. Los experimentos realizados en México por Zúñiga Molina y Gaviria (2010) sugieren que el porcentaje promedio de probable copia o trampa en todos los grados entre 2006 y 2009 fue alto—ascendiendo de 4,5% en 2006 al 7% en 2008 (OCDE, 2011, p. 61). Además, un Equipo de Revisión de la Organización de Cooperación y Desarrollo Económico (OCDE) observó que en muchas de las escuelas que visitaron, algunos

profesores pidieron a sus alumnos de bajo rendimiento que no asistieran a clase el día que se administraban las evaluaciones de ENLACE, mientras que otros ayudaron activamente a los estudiantes a completar el examen (OCDE, 2012, p. 126). En 2014, la Secretaría de Educación Pública anunció la suspensión de los exámenes ENLACE, citando numerosos casos en los que los profesores les dieron a sus alumnos las respuestas de la prueba de antemano, o corrigieron las respuestas incorrectas con el fin de mejorar las puntuaciones obtenidas (Martínez Carballo, 2014). La OCDE lo atribuye al hecho de que los maestros se beneficiarían con bonificaciones salariales si sus estudiantes obtuviesen un buen desempeño en el examen, y que el examen era utilizado a menudo por los medios de comunicación para hacer rankings de escuelas (OCDE, 2011, p. 85).

A pesar de las limitaciones que las pruebas censales enfrentan a la hora de permitir análisis más profundos de los factores asociados, la experiencia indica que en realidad los datos de factores “más simples” procedentes del IDEB y del ISCE están siendo utilizados. Por el contrario, los factores asociados más detallados, tales como los producidos por las evaluaciones muestrales de las escuelas primarias de Guatemala¹⁹, o los datos sobre factores de submuestras extraídos de Prova Brasil, no están teniendo el impacto deseado en términos de informar a la política, o de motivar la introducción de cambios para impulsar una mejora en la calidad de la educación. En síntesis, se observa una tendencia a utilizar datos de contexto más limitados para los índices de alto impacto, mientras que los análisis más profundos, aunque más importantes y potencialmente más útiles, están siendo sub-utilizados y reservados para las audiencias académicas, en el mejor de los casos.

El uso y la difusión de los datos también se puede observar en la manera en la cual las Unidades de Evaluación hacen pública la información a través de sus sitios web. Aquí el escenario es bastante heterogéneo entre los distintos países, tal vez no tanto en términos de qué y cómo se publica digitalmente, sino más bien en la manera en la que se organizan esos sitios web. En cuanto a la publicación de datos, se observan al menos ocho maneras diferentes de informar los resultados y de explicar cómo interpretarlos²⁰: (i) Informes completos con los resultados nacionales y sub-nacionales para la aplicación de cada evaluación; (ii) Informes que muestran el progreso longitudinal en los últimos 5-10 años; (iii) Informes individuales para los estudiantes y las escuelas en el caso de las evaluaciones de altas consecuencias

(generalmente protegidos con contraseña para garantizar la confidencialidad); (iv) Bases de datos para uso académico (investigación); (v) Análisis secundarios llevados a cabo por las unidades de evaluación o por investigadores externos; (vi) Recomendaciones pedagógicas basadas en ítems liberados y resultados de evaluaciones; (vii) Motores de búsqueda en línea para encontrar resultados (cuando éstos no son confidenciales) y para encontrar ejemplos de los contenidos curriculares evaluados; y (viii) Variedad de presentaciones audiovisuales, incluyendo PowerPoint, Prezi y videos.

A pesar de la variedad de informes que se encuentran en formato digital, es necesario resaltar que los sitios web oficiales a menudo no están bien diseñados: se necesita tiempo y esfuerzo para poder encontrar informes, para distinguir los informes según los diferentes tipos de evaluaciones, y muchas veces para abrir o descargar archivos debido a fallas del sistema. Tanto la información general como los datos específicos a menudo se repiten en diferentes sub-sitios, que se encuentran bajo nombres de enlace confusos, o publicados todos juntos sin una especificación clara de las distintas audiencias. Estos defectos pueden no ser tan importantes para los investigadores, que tienden a tener la voluntad y la experiencia para llevar a cabo búsquedas complejas. Sin embargo, si se los juzga como instrumentos para usuarios no expertos, tales como profesores o ciudadanos interesados, estos sitios web todavía necesitan mejorar para ser más accesibles y fáciles de consultar.²¹ En apariencia, es como si las autoridades de las unidades de evaluación les ordenasen a los diseñadores de sitios web que carguen toda la información disponible, sin realizar ninguna supervisión sobre la forma en que la información se presenta a los usuarios.

Nuestra revisión encontró que los sitios web oficiales no ofrecen datos sobre cuántos de los informes que son publicados digitalmente se encuentran disponibles al público en versiones impresas o si son o no reenviados por correo o entregados en el marco de talleres de formación docente o en eventos similares.

Independientemente de cómo los informes son entregados a las audiencias, ya sea en formato digital o en papel, lo que sigue siendo un problema importante es la falta de capacidad por parte de las comunidades educativas—especialmente al nivel de escuela—para entender los datos que se presentan en formatos textuales que les resultan ajenos, y para los cuales no han sido preparados durante sus años de formación como futuros educadores.

Los formatos de texto a los cuales hacemos referencia son aquellos que incluyen datos en una variedad de presentaciones gráficas, incluyendo cuadros, barras, escalas y tablas comparativas, entre otros. Los países han intentado diferentes maneras de presentar estos formatos de un modo más simple, recurriendo a todo tipo de dibujos, fotografías, referencias de color, así como al uso de lenguaje simplificado y una profusión de ejemplos, tanto de manera estática (en papel) o dinámica (como en las presentaciones tipo Prezi y videos).

La mayoría de los objetivos nacionales de educación incluyen los resultados de aprendizajes como un indicador clave para medir la calidad.

Sin embargo, por varias razones, no existe un formato que parezca dar resultados satisfactorios en cuanto a la interpretación y el uso de estos informes.²² Por lo menos dos factores posibles pueden estar obstaculizando el progreso en este sentido. En primer lugar, como se mencionó antes, los educadores locales (profesores, directores, supervisores), así como las familias, parecen no tener suficientes habilidades de lectura y de matemática básica para interpretar textos discontinuos, no verbales. Ni siquiera los talleres presenciales parecen ser suficientes para contrarrestar esta deficiencia en la preparación cognitiva. En segundo lugar, se puede suponer que las evaluaciones de bajas consecuencias ('low-stakes'), es decir, aquellas cuyos resultados no tienen consecuencias directas para los individuos, proporcionan poco incentivo para hacer el esfuerzo y comprender el significado y la relevancia de estos textos. Sin embargo, éstas son las impresiones de los autores, basadas en su experiencia propia y en las entrevistas realizadas. Para entender este problema y encontrar soluciones adecuadas se necesitaría más investigación sobre estrategias adecuadas de comunicación.

Por último, consideramos el impacto que tienen los datos de las evaluaciones en la opinión pública y en la sociedad en la

actualidad, el cual se encuentra considerablemente mediado por la prensa. Éste es probablemente uno de los principales signos de evolución en el campo de las evaluaciones a gran escala en América Latina. Todos los entrevistados que fueron consultados, además del conocimiento que tienen los autores sobre el escenario de evaluación de aprendizajes en la región, coinciden en sus impresiones acerca de cómo los resultados de evaluación han ayudado a instalar el tema de la calidad educativa—entendida en términos de resultados de aprendizajes—en los debates nacionales y transnacionales en América Latina.

La mayoría de los objetivos nacionales de educación, tal como se expresa en una variedad de planes gubernamentales a mediano y largo plazo, así como las directrices de agencias internacionales y las medidas de apoyo para mejorar la calidad educativa en la región, incluyen los resultados de aprendizajes como un indicador clave para medir la calidad de la educación y el progreso a través del tiempo.

Hace ya una década, la prensa había empezado a prestar atención a estos indicadores, sobre todo después de que los países latinoamericanos comenzaran a participar en PISA (Programme for International Student Assessment) y se puso de manifiesto qué tan retrasados están los países de la región en términos de sus resultados de aprendizaje. Inicialmente, se recibían los resultados comparativos internacionales con gran expectativa, sobre todo con un interés por publicar noticias negativas, demonizar a los sistemas educativos a través de los titulares, los cuales se enfocaban únicamente en rankings, y nunca en un tratamiento más profundo de las causas o de la complejidad que implica una mejora de las condiciones educativas. Algo similar ocurría con los resultados nacionales, especialmente en aquellos países que estaban implementando, o que comenzaban a implementar, evaluaciones censales en los últimos años de la educación secundaria.

Hoy en día, a pesar de que las noticias rara vez son lo suficientemente buenas en términos 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. 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. También están viendo la necesidad de aplicar medidas más coherentes que conduzcan a una mejora de estos resultados. Un contragolpe en este sentido también puede ocurrir, si es que ya no está ocurriendo, debido a que los políticos tienden a tomar decisiones apresuradas con el fin de demostrar que se están tomando medidas para cambiar la situación. Por lo tanto, la nueva tendencia a utilizar programas de incentivos para escuelas y docentes que buscan mejorar los resultados de estas evaluaciones podría tener realmente un impacto negativo si estos programas no se planifican y se ejecutan luego de una cuidadosa reflexión y estudio de políticas similares en contextos nacionales comparables.

PARTICIPACIÓN EN PRUEBAS INTERNACIONALES

Conceptualmente, se podría argumentar que las evaluaciones internacionales no deben ser tratadas como una categoría diferente de las evaluaciones nacionales. Sin embargo, debido al hecho de que estas evaluaciones son diseñadas e implementadas por organismos supranacionales, su connotación e implicaciones, tanto en términos técnicos como políticos, merecen ser tratados en una sección separada.²³

Durante el período considerado, los países latinoamericanos han seguido participando en evaluaciones internacionales, con algunas variaciones en el tiempo con respecto a las evaluaciones en las cuales deciden participar, y también en el número de países que participan en cada una de las diferentes evaluaciones (Ver **Anexo 4**). Las tendencias muestran que a) un número creciente de países de la región ha decidido participar en PISA y LLECE (Laboratorio Latinoamericano de la UNESCO para la Evaluación de la Calidad de la Educación); b) sólo entre dos y tres países han participado en TIMSS entre 1999 y 2011, c) los países latinoamericanos no muestran mucho interés en PIRLS²⁴ (un solo país participó en la última aplicación, mientras que el número de países a nivel mundial ha aumentado); y d) más países de América Latina se han interesado en la evaluación de competencias ciudadanas a través de la evaluación ICCS de la IEA (antes llamada CIV-ED).

Las razones para participar en evaluaciones internacionales, de acuerdo con los entrevistados, han sido prácticamente las mismas que las que se esgrimían en el año 2003. En primer lugar, los países consideran que es importante y necesario compararse tanto con sus “pares” regionales,

como es en el caso del LLECE, como con países altamente industrializados de Asia, Europa y América del Norte, principalmente a través de las evaluaciones PISA de la OCDE. En segundo lugar, las unidades nacionales de evaluación han encontrado, y todavía encuentran en PISA una gran oportunidad para el desarrollo y el fortalecimiento de sus capacidades técnicas y metodológicas. A pesar de que los países latinoamericanos no tienen un papel significativo en el diseño y el desarrollo de esta prueba, el proceso de participación en esos programas ofrece múltiples oportunidades para el intercambio técnico y para la adquisición de conocimiento de metodologías de medición cada vez más complejas. Ésta fue también una fuerte motivación para unirse a la primera evaluación LLECE en 1997. Hoy en día, sin embargo, esa percepción ha cambiado ya que LLECE ya no parece proporcionar ese tipo de incentivo. La participación en LLECE, sin embargo, no ha disminuido. Una investigación ulterior podría analizar las razones de esta continuidad.

Una de las objeciones más importantes a participar en PISA, aún hoy, es que no tiene sentido invertir tiempo y recursos económicos en un proyecto de evaluación donde los países latinoamericanos siempre salen en los últimos puestos del ranking. Estas objeciones, sin embargo, se enfrentan a fuertes argumentos en contra por lo menos en dos frentes. Por un lado, se ha observado una cierta mejora de los resultados, como es el caso de Perú, lo que sugiere que una combinación de mejores políticas educativas, más inversión y mejor desempeño económico en general se traduce en una mejora de los resultados. En segundo lugar, PISA ofrece una aplicación regular cada tres años—en teoría con instrumentos mucho más fiables—lo cual es altamente valorado por los países interesados en hacer análisis longitudinales de progreso en la calidad educativa. Además, también se ha argumentado que salir últimos en el ranking puede tener efectos positivos para estimular el debate nacional y la obtención de apoyo público para introducir medidas de política difíciles o controversiales, como las reformas en la formación docente o en los programas de evaluación de maestros.

Tal como se desprende de las opiniones que los entrevistados tienen sobre PISA, en algunos países como México, Brasil y Chile, la participación en la prueba es parte de un interés geopolítico mayor—ya sea porque son Estados-miembro de la OCDE (y, por tanto, están obligados a ser evaluados por PISA), o porque les resulta atractivo en términos estratégicos y económicos formar parte del “entorno” de la OCDE. Esto podría explicar por qué algunos países como Argentina todavía participan en PISA, incluso

cuando se descalifica y descarta el uso de los datos, o por qué México permite que la OCDE no sólo evalúe el aprendizaje a través de PISA, sino que también discuta recomendaciones y directrices específicas para la política educativa nacional.

En general, sin embargo, como es también el caso de las evaluaciones nacionales con bajas consecuencias, existe la percepción de que las evaluaciones internacionales proporcionan una gran cantidad de información que no es suficientemente utilizada para realizar análisis más profundos de los factores contextuales que explican los resultados, o para comparar con rigor el desempeño de un país con el de los sistemas educativos de todo el mundo para obtener aportes de información a las reformas o mejoras educativas. En este sentido, Finlandia se ha convertido en el paradigma de la “historia exitosa”, generando interés en toda América Latina en lugar de la descalificación por ser un país altamente desarrollado (y por lo tanto poco adecuado para comparaciones válidas con América Latina).

A pesar de la sub-utilización real o supuesta de los datos, se observa cómo el marco de evaluación, los constructos y la estructura de PISA han influido poderosamente los modelos y los contenidos de las evaluaciones nacionales de América Latina. Simultáneamente, también han influido en el diseño curricular, tanto en su contenido como en su estructura. Ésta es quizás una conclusión ambiciosa que se desprende de evidencia insuficiente en el contexto de este informe. Sin embargo, los autores han podido observar esta tendencia en una serie de casos nacionales a través de estudios previos y de trabajos de consultoría en diferentes países. Sólo para proporcionar un ejemplo, fue después de PISA que las competencias lectoras, o la comprensión lectora para ser exactos, llegó a ser definida casi de manera universal en tres procesos cognitivos distintos, aunque complementarios: localizar información en un texto, hacer inferencias y evaluar la calidad del texto para transmitir significado de manera adecuada. Antes de PISA, uno podía encontrar definiciones de “lectura” diferentes para cada plan de estudio o cada prueba que se estuviese aplicando. Esto ha cambiado.²⁵

Por último, y de nuevo en referencia a PISA—claramente la evaluación internacional más influyente en la región en la actualidad—vale la pena mencionar que una nueva versión de la prueba, diseñada específicamente para los países en desarrollo, será implementada en breve. Se llamará PISA para el Desarrollo (PISA-D)—en inglés PISA for Development, PFD. Tres países latinoamericanos (Ecuador, Guatemala

y Paraguay) han confirmado su participación. PISA-D tiene un objetivo doble. En primer lugar, incluirá ítems que miden habilidades muy básicas, que a su vez ayudarán a discriminar mejor y revelar diferencias en el rendimiento entre los millones de estudiantes que muy probablemente estarán en el nivel “por debajo de 1” de la versión normal de PISA. En tanto las habilidades más básicas sean medidas y comparadas, habrá más posibilidades de análisis para comprender los aspectos más sutiles de aprendizaje dentro de estas poblaciones y por lo tanto se podrá obtener información más relevante para las políticas de mejora. El segundo objetivo será el de darles a los países con unidades de evaluación débiles o de escasa experiencia la posibilidad de desarrollar la capacidad técnica hasta que puedan, si así lo desean, participar en el programa regular de evaluación global de PISA. Es de acuerdo con este segundo objetivo que PISA-D planea emparejar los equipos de los países inexpertos con los técnicos más fuertes en la región para promover el intercambio, la colaboración y el desarrollo técnico. El hecho de que tres de los cinco países que han confirmado inicialmente su participación en PISA-D estén en América Latina demuestra el alto nivel de interés que PISA genera en la región.

Lecciones aprendidas

Nuestro análisis encontró que ha ocurrido un progreso claro en la última década. Si bien reconocemos que sólo hemos cubierto tres de las siete dimensiones originales en Ferrer (2006) y que nuestro informe tiene un alcance limitado, en general, creemos que es justo decir que América Latina ha podido mantener un proceso de construcción de sistemas de evaluación educativa. Mientras que los niveles de mejora pueden variar entre los distintos países, no se encontraron signos de retroceso. Por el contrario, hay muchas señales de que los sistemas han madurado.

Sin embargo, no todas las noticias son buenas. Nuestro análisis encontró que, a pesar de los progresos observados, aún existen muchas debilidades. Si bien esto puede haber sido evidente para los expertos todo este tiempo, la experiencia de los países de América Latina vuelve a confirmar que la construcción de un sistema de evaluación educativa efectivo que informa y apoya las mejoras en la calidad de la educación es una tarea laboriosa y compleja que requiere una alineación de los objetivos de aprendizaje, de los estándares, del currículo y de otros aspectos fundamentales como la formación

docente (tanto antes como durante el ejercicio de la profesión). Esto es, de hecho, una de las tres dimensiones del marco de la Evaluación de Estudiantes SABER (Clarke, 2012). Crear capacidad y adaptación de herramientas y procesos para responder a las cambiantes necesidades de información del sistema educativo exige esfuerzos. Podría decirse que entre los cinco países que consideramos, sólo Brasil y Colombia están llegando al punto en que podemos afirmar que han desarrollado sistemas de evaluación que articulan la producción y uso de los datos con usuarios que dentro y fuera del sistema escolar. Creemos que ésta es un área en la que se necesita más investigación para documentar mejor los avances y los obstáculos que los países de América Latina están enfrentando para lograr un nivel avanzado en la alineación del sistema.

Estas son lecciones importantes para los países de otras regiones que están menos avanzados en el desarrollo de sistemas de evaluación de aprendizajes (muchos de los cuales experimentan situaciones similares a las que se vivían en los países latinoamericanos hace dos décadas) y para los donantes que los apoyan. Deben ser realistas sobre lo que puede lograrse en períodos de tiempo relativamente cortos, y persistentes en sus esfuerzos de desarrollo de capacidades. La falta de realismo puede conducir a expectativas frustradas, y la falta de consistencia en el desarrollo de capacidades puede, ciertamente, conducir a resultados pobres en los sistemas de evaluación.

La elección entre poner institutos autónomos a cargo de la evaluación o mantener esas funciones en unidades especializadas dentro de los MdE es menos evidente de lo que mucha gente habría pensado hace 10 años. Entre los casos que estudiamos, las unidades (en Guatemala y Perú) han sido capaces de hacer un sólido trabajo en general. Aunque, naturalmente, no sabemos cuál habría sido la situación si hubiesen seguido el camino alternativo de establecer institutos autónomos, el hecho de que las funciones de evaluación se encuentren dentro del MdE no es necesariamente una receta para el fracaso.

Ahora bien, existe evidencia que sugiere que los institutos tienen algunas ventajas frente a las unidades de los MdE que derivan de su mayor nivel de autonomía. En primer lugar, en los lugares donde los obstáculos burocráticos (por ejemplo, las reglas de contratación) hacen que sea difícil mantener la presencia de personal técnico altamente calificado dentro del Ministerio, la autonomía puede ayudar a crear mejores condiciones para desarrollar capacidades. El ICFES y el INEP pueden haberse beneficiado en este

aspecto. En segundo lugar, si se mantiene, la autonomía puede proporcionar un mayor nivel de protección contra las presiones políticas que las que enfrentan las unidades que responden directamente a las autoridades políticas en el MdE.

Vale la pena realizar dos advertencias con respecto a la caracterización anterior. En primer lugar, la autonomía legal no garantiza la autonomía política: las presiones políticas pueden debilitar la autodeterminación de órganos aparentemente autónomos. En segundo lugar, la autonomía no resuelve, por sí misma, el problema de la escasa credibilidad que azota a muchos sistemas de evaluación. A menos que los mecanismos de gobierno de los institutos de evaluación autónomos ofrezcan una supervisión adecuada de sus actividades, las dudas sobre la fiabilidad de sus resultados pueden permanecer. La lección para los países que están considerando la creación de institutos de evaluación autónoma es que éstos exigen un fuerte marco legal—incluyendo el establecimiento de acuerdos de gobierno que garanticen una supervisión técnica adecuada, financiamiento y la capacidad de tomar decisiones sobre recursos humanos necesarios de manera independiente.

La transformación de los datos generados a través de las evaluaciones de aprendizaje en información que sea útil y utilizable por diversos actores del sistema sigue siendo un desafío. En parte, esto se debe a las continuas debilidades en cómo se utilizan y difunden los resultados. En algunos casos puede haber problemas en cuanto a la forma en la que los resultados se comunican (por ejemplo, demasiados datos o sitios web complejos que son difíciles de consultar).

Pero la presentación de informes es sólo una parte del problema. La demanda efectiva de datos y la capacidad para analizarlos, también es en muchos casos una restricción vinculante. Lo cierto es que los países de América Latina siguen luchando para conseguir que profesores y familias se interesen en los datos generados por los sistemas de evaluación de aprendizajes. Ésta es un área fértil para la investigación con un enfoque experimental que pruebe formas alternativas de comunicar los resultados e involucrar a los actores del sistema educativo en el uso de esos resultados.²⁶

Quizás la novedad más interesante de cómo los resultados de evaluación se pueden comunicar de manera sencilla y práctica es utilizando índices, como en el caso de Brasil y Colombia. Los índices no sólo simplifican las

presentaciones de los resultados sino que también pueden proporcionar una perspectiva más amplia, que no se base en presentar los resultados del aprendizaje como una medida única de la calidad.

Por otro lado, la sociedad civil y los medios de comunicación en muchos países han demostrado un aumento de la capacidad analítica para utilizar los datos. El desarrollo de su capacidad es, probablemente, una tarea más directa que el desarrollo de las capacidades de las organizaciones estatales que tienen más requisitos burocráticos y políticos. El desarrollo de estas capacidades en los medios de comunicación y en la sociedad en América Latina ha sido, probablemente,

Para aquellos países que están empezando a construir sus sistemas de evaluación educativa, la lección es que éste es un camino que vale la pena recorrer, aun cuando exige un compromiso a largo plazo y un apoyo político sostenido.

parte del (lento) proceso de ejercer los derechos democráticos y, por tanto, difícil de acelerar y gestionar a través de esfuerzos externos. Sin embargo, si se piensa en los países en los que los donantes han participado activamente en el fortalecimiento de las capacidades de las organizaciones de la sociedad civil y los medios de comunicación, vale la pena explorar los esfuerzos dirigidos a la creación de capacidades para exigir y analizar datos sobre educación.

Los países de América Latina están participando cada vez más en las pruebas internacionales. Esto es, por

ahora, una tendencia bien establecida. Aparentemente, desde el punto de vista de quienes toman las decisiones, los beneficios de la participación exceden los costos, incluyendo los costos políticos de ser identificados como participantes con desempeños relativamente pobres.

La experiencia indica que esos beneficios son tanto técnicos como políticos. En el aspecto técnico, la participación en las evaluaciones internacionales de alta calidad tales como PISA es visto como una actividad de inversión: el desarrollo de capacidades resultante proporciona un fuerte incentivo para que un número importante y creciente de países de América Latina participen en esas pruebas. PISA para el Desarrollo puede contribuir aún más a esta tendencia.

En el aspecto político, hay algunos indicios de que las autoridades educativas en América Latina están comenzando a valorar la oportunidad de comparar sus resultados con los de otros países. Si bien en el pasado la evaluación comparativa regional puede haber sido más segura y más cómoda, ya que varios países de la región son miembros de la OCDE o aspiran a serlo en un futuro no muy lejano (por ejemplo, Colombia o Perú), el valor político de participar en evaluaciones regionales puede disminuir con respecto a los internacionales a través del tiempo.

Una revisión de la experiencia de más países de América Latina, y un mayor número de entrevistas, probablemente habría permitido profundizar en los factores que ayudan y que limitan el desarrollo de los sistemas de evaluación de aprendizajes. Sin embargo, las tendencias generales que hemos identificado en nuestro análisis proporcionan algunas lecciones valiosas tanto para los países de América Latina como para otras regiones del mundo en vías de desarrollo. La evaluación de la educación se está consolidando en la región y hay claros signos de progreso aunque sigue siendo necesaria una mayor consolidación. Hacer buen uso de los datos para mejorar la calidad de la educación sigue siendo el talón de Aquiles de estos esfuerzos. Como aspecto positivo, rescatamos que existen buenas prácticas en varios países de los cuales toda la región puede aprender. Y para aquellos países (en América Latina y en otras partes del mundo en desarrollo) que están empezando a construir sus sistemas de evaluación educativa, la lección es que éste es un camino que vale la pena recorrer, aun cuando exige un compromiso a largo plazo y un apoyo político sostenido.

Notas

1. A partir de finales de los 90, los países de América Latina experimentan una ola de reformas educativas que fueron más allá del tradicional y exclusivo enfoque en la ampliación de la cobertura y trataron de mejorar la calidad de varias maneras. Esa ola creó la demanda inicial de sistemas de evaluación de aprendizajes. Una lista completa de los programas de evaluación está incluida en el **Anexo 3**.

2. Los informes de los países SABER (<http://saber.worldbank.org/index.cfm?indx=9&pd=5>, consultados el 7 de junio de 2015) utilizan una metodología común para evaluar los sistemas de evaluación de los estudiantes en los países en desarrollo. (SABER, por sus siglas en inglés, es un enfoque sistémico del Banco Mundial para lograr mejores resultados en la educación). En este trabajo se considera dos de las tres dimensiones del marco conceptual SABER para Sistemas de Evaluación de Estudiantes (Clarke, 2012).

3. Las dimensiones en el estudio original no cubiertas en este informe son: Currículos y Estándares, Instrumentos, Informes y Sistemas Subnacionales. Dada la naturaleza de este trabajo -basado en entrevistas y revisiones de documentos disponibles en línea, todo esto realizado en un corto período de tiempo- fue particularmente difícil analizar estas dimensiones ya que hacerlo hubiese requerido visitas de campo (por ejemplo, para recopilar datos sobre los sistemas subnacionales) o la recopilación y revisión de documentos que no siempre se encuentran disponibles en línea (por ejemplo, instrumentos).

4. No tenemos conocimiento de ningún estudio que mida los costes de creación y funcionamiento de los sistemas de evaluación de aprendizajes en la región más allá de un estudio de PREAL (Wolff, 2007). Sin embargo, como una referencia, el presupuesto del INEE de México aumentó de 300 millones de pesos en 2013 a \$ 613 millones de pesos en el año 2014 y \$ 1.000 millones de pesos en 2015 (o US\$ 69 millones). En comparación, el ya bien establecido INEP en Brasil tuvo un presupuesto de \$ 854 millones de reales en 2013 (el equivalente a US\$ 393 millones). En términos per cápita, el presupuesto del INEP es casi 4 veces más grande que el del INEE.

5. Una lista de los sitios web se incluye en la sección de referencias.

6. Ver **Anexos 1 y 2** para el protocolo de entrevista y la lista de las personas entrevistadas.

7. En este informe, vamos a utilizar la expresión “unidades de evaluación” para referirnos a esas oficinas o divisiones dentro de los Ministerios de Educación (MdE) dedicados a la tarea específica de medición de logros de aprendizaje. Cuando nos referimos a los Institutos, será en referencia a las organizaciones autónomas o semi-autónomas fuera de los MdE, que tienen la misma función de medir logros de aprendizaje. En ningún caso estamos asumiendo que las funciones de evaluación de aprendizaje en realidad se han convertido en un sistema, ya que eso implicaría que los programas de evaluación han logrado articularse verdaderamente con otros componentes de la administración educativa, tales como el desarrollo curricular, la formación del profesorado, y otras políticas y programas relacionados.

8. En SABER, el contexto facilitador (“Enabling Context” en inglés) cubre áreas tales como el marco legislativo o políticas para las actividades de evaluación; liderazgo que rodea la actividad de evaluación; compromiso público con la actividad de evaluación; los arreglos institucionales para el diseño, la implementación, o el uso de los resultados de la actividad de evaluación; la disponibilidad de fuentes de financiación suficientes y estables y la presencia de agentes evaluadores competentes y de maestros (Clarke, 2012).

9. A pesar de que la unidad se conoce como DIGEDUCA, el nombre legal bajo el cual se estableció es la Dirección General de Evaluación, Investigación y Estándares Educativos.

10. Bajo el nuevo marco legal, el INEE es responsable de coordinar el Sistema Nacional de Evaluación de la Educación, de regular todos los componentes de ese sistema, de evaluar el sistema y expedir los lineamientos de política sobre la base de resultados de evaluación. En este trabajo se considera solamente la evaluación del aprendizaje de los alumnos y no las otras funciones de evaluación de las cuales el INEE es responsable.

11. En América Latina, las instituciones públicas a menudo se definen como autárquicas, autónomas o semiautónomas. La diferencia no siempre es clara, aunque este regida por definiciones sociológicas, por lo que en este trabajo los autores han respetado la terminología utilizada oficialmente por las instituciones estudiadas.

Notas, cont.

12. La SEP en realidad puede objetar las resoluciones del INEE, siempre que las explicaciones pertinentes y alternativas se presenten a las autoridades nacionales. La reforma educativa de 2012 en México, declaró al Instituto Nacional de Evaluación Educativa (INEE) como una “entidad pública autónoma, con personalidad jurídica y patrimonio propio”, con el fin de garantizar la prestación de “servicios educativos de alta calidad” en el país, y garantizar la producción de datos bien fundados, pertinentes y fiables sobre el sistema educativo (INEE, 2015, p. 41). El INEE tiene autoridad en materia de evaluación educativa, y es responsable de “la evaluación de la calidad, el rendimiento y los resultados del sistema educativo nacional” (ibíd.). Por lo tanto, su deber es emitir las directrices que las autoridades educativas federales y locales deben seguir a la hora de realizar sus funciones de evaluación. Mientras tanto, las autoridades federales y locales están obligados a cumplir con estas directrices y a utilizarlas para desarrollar políticas educativas basadas en los resultados de estas evaluaciones. En otras palabras, el INEE no tiene una relación de subordinación con respecto a la Secretaría de Educación Pública, sino más bien uno de “coordinación y colaboración” (Bracho y Zorrilla 2014).

13. Durante la última década, los sindicatos se han vuelto menos activos en términos de oposición a las evaluaciones de aprendizajes de los estudiantes, probablemente debido al hecho de que este tipo de evaluaciones han demostrado no ser diseñadas como instrumentos de castigo contra los maestros. Sin embargo, se han hecho más activos, aunque no necesariamente con éxito, en la prevención de la ejecución de los programas de evaluación de docentes diseñados para proporcionar incentivos económicos o para regular el acceso a nuevos cargos docentes en el sistema educativo.

14. Incluso las pruebas con bajas consecuencias “low-stakes” tienden a ser más regulares en cuanto a las fechas de administración que hace una década. Sin embargo, las pruebas con altas consecuencias “high-stakes”, debido a su naturaleza (es decir, debido a que tienen consecuencias para los individuos), tienen que ser más rigurosas en este sentido.

15. Ambas evaluaciones, EXCALE del INEE y Enlace de la SEP se han suspendido hasta que un nuevo programa de evaluación que involucra ambas instituciones sea implementado. El programa será responsabilidad del INEE,

como parte de su nuevo mandato legal para coordinar un sistema nacional de evaluación, y será llamado PLANEA (Plan Nacional de Evaluación de Aprendizajes).

16. Por formato nos referimos a la forma de presentar los resultados de acuerdo con la metodología utilizada para la medición de logros de aprendizaje.

17. “Factores asociados”, es la expresión utilizada en español para definir o describir el análisis de datos del contexto utilizados para explicar estadísticamente los resultados de aprendizaje. Los análisis suelen incluir datos del contexto escolar (por ejemplo, licencias de los maestros, el número de libros en la biblioteca, acceso a Internet), así como características de los estudiantes y de sus familias (por ejemplo, número de libros en el hogar, el nivel de educación de los padres, los ingresos, el acceso a internet y TV por cable). Cada uno de estos datos se denomina “factor”. En términos más generales, los factores pueden ser entendidos como datos de fondo o de contexto utilizados para explicar los resultados de aprendizaje.

18. IDEB no controla el nivel socioeconómico o el tamaño de la escuela.

19. En los últimos años Guatemala ha mostrado un mayor interés en hacer un uso más eficiente de los resultados de las evaluaciones. En 2013, a través de paneles de discusión, entrevistas con actores claves en el campo de la educación y consultas con expertos, el Ministerio de Educación (MINEDUC) debatió la necesidad de utilizar las evaluaciones como criterios para cosas tales como clasificación de los estudiantes y aumentos salariales para docente, entre otros (MINEDUC, 2013, p. 33). MINEDUC reconoció las deficiencias en las estrategias de comunicación y difusión de los resultados, y desarrolló los informes “EVALUAR” para los maestros, que resumen los resultados de evaluaciones y ofrecen recomendaciones para hacer un mejor uso de esos resultados (MINEDUC 2015). MINEDUC también le proporcionó a los directores de escuela y maestros manuales de matemática y lenguaje que contienen estrategias para mejorar la calidad de la enseñanza.

20. Muchos de estos informes incluyen datos de evaluación de aprendizajes desglosados por sexo, origen étnico o lingüístico, entre otros. El número y la variedad de los informes disponibles en los sitios web, sin embargo, requeriría más tiempo para el análisis si se pretendiese proporcionar una explicación completa de todas las

Notas, cont.

variables que se utilizaron durante la recolección de datos, para luego informar sobre los resultados de aprendizaje y sus factores asociados. Perú ofrece un ejemplo útil. Desde 2007, el Ministerio de Educación administra una evaluación censal del rendimiento de los estudiantes (ECE) en sus escuelas. ECE evalúa a los estudiantes de segundo grado en matemáticas y comprensión lectora. También evalúa a los estudiantes de cuarto grado cuyo primer idioma es una lengua indígena y que asisten a escuelas clasificadas como de Educación Intercultural Bilingüe (EIB), en la comprensión lectora en español como segunda lengua. Además, en 2007, 2008, 2010 y 2012, la evaluación también probó la comprensión lectora de los alumnos en uno de cuatro idiomas nativos—Awajún, Quechua Cusco–Collao, Aimara y Shipibo-Conibo, dependiendo de la región donde se administró. En 2014, la prueba también se ofreció por primera vez en Quechua Chanka. A pesar de mostrar una mejora con respecto a 2012, en el 2014 los estudiantes que rindieron el examen demostraron mala comprensión lectora en ambos idiomas (MINEDU 2014 y MINEDU 2012).

21. Por el contrario, se podría pensar en aplicaciones para tablets o smartphones en las cuales las opciones de navegación tienden a ser más limitadas, pero son a su vez más claras y más fáciles de usar.

22. Se observa un caso interesante en Uruguay, donde la Administración Nacional de Educación Pública administra evaluaciones en línea que utilizan ordenadores proporcionados por el Plan Ceibal. Los maestros pueden monitorear a sus alumnos desde sus propias computadoras mientras que realizan el examen, tener acceso a los resultados y ver las puntuaciones de cada estudiante, así como sus respuestas individuales. Los resultados de las pruebas vienen con información para ayudar a los profesores a interpretarlos. Por ejemplo, una descripción detallada de un ítem para comprender mejor una pregunta, analizar errores conceptuales persistentes, y planificar intervenciones correctivas (Luaces, 2014). Los maestros también tienen acceso a “guías didácticas”, desarrolladas por especialistas para ayudarlos a mejorar sus metodologías de enseñanza.

23. Vale la pena mencionar que PREAL publicó en 2003 los resultados de un estudio regional (Ferrer y Arregui, 2003) con el apoyo financiero del Consorcio Peruano de Investigación Económica y Social (CIES). Este estudio también puede ser tomado como una referencia en el

tiempo con respecto a los cambios y continuidades en la participación de América Latina en las evaluaciones internacionales a lo largo de una década.

24. Los entrevistados no ofrecieron ninguna explicación para esta falta de interés. Es posible que las autoridades educativas de la región consideren que el costo de participar en tantas pruebas internacionales es simplemente demasiado alto.

25. El isomorfismo en este campo tiene varios detractores y, posiblemente, podría ser una razón más para oponerse a las evaluaciones internacionales como una nueva forma de colonialismo cultural. Sin embargo, otras voces argumentan que la pedagogía y el currículo están plagados de terminología y definiciones erráticas y arbitrarias, y por lo tanto debe ser una buena noticia que se haya alcanzado un acuerdo conceptual y semántico. El debate, sin embargo, sigue abierto y probablemente continuará durante muchos años más.

26. Para un ejemplo reciente de México, ver Hoyos, García-Moreno y Patrinos (2015).

Referencias

Bracho, Teresa y Margarita Zorrilla (2014). El INEE en la reforma educativa: De su naturaleza jurídica y sus atribuciones.

CIEN y Diálogo Interamericano (2015). El estado de las políticas públicas docentes: Guatemala. Ciudad de Guatemala.

Clarke, Marguerite (2012): What Matters Most for Student Assessment Systems: A Framework Paper. Washington, DC: The World Bank. (Disponible en: <http://documents.worldbank.org/curated/en/2012/04/16238771/matters-most-student-assessment-systems-framework-paper>)

de Hoyos, Rafael, Vicente A. García-Moreno y Harry A. Patrinos (2015): The impact of low-stakes accountability intervention with diagnostic feedback: Evidence from México, (mimeo). Washington, DC: The World Bank.

Ferrer, Guillermo y Patricia Arregui (2003): Las pruebas internacionales de aprendizaje en América latina y su impacto en la calidad de la educación: Criterios para guiar futuras aplicaciones, Documento No. 26. Santiago de Chile: PREAL.

ESCALE (2015). "Magnitudes". [Base de Datos]. Lima: Ministerio de Educación del Perú. (Disponible en: <http://escale.minedu.gob.pe/magnitudes>)

Ferrer, Guillermo (2006): Educational Assessment Systems in Latin America: Current Practices and Future Challenges. Washington, DC: PREAL.

Fortín, Álvaro (2013). Evaluación Educativa Estandarizada en Guatemala: Un camino recorrido, un camino por recorrer. Ciudad de Guatemala: Ministerio de Educación. (Disponible en: https://www.mineduc.gob.gt/PORTAL/contenido/enlaces/documents/Documento_Evaluacion.pdf)

INEE (2015): Reforma Educativa: Marco Normativo. Ciudad de México: INEE.

Luaces, Oscar (2014). Evaluaciones en línea: Una innovación educativa en el marco del programa de una computadora por alumno. Montevideo: Administración Nacional de Educación Pública. (Disponible en: http://www.anep.edu.uy/sea/wp-content/uploads/2014/12/Evaluaciones-en-l%C3%ADnea_-para-web.pdf)

Martínez Carballo, Nurit (2014, Feb 4). "SEP: En ENLACE hubo corrupción". El Universal. (Disponible en: <http://archivo.eluniversal.com.mx/nacion-mexico/2014/impreso/sep-en-enlace-hubo-corrupcion-212806.html>)

MINEDU (2012). Resultados de la Evaluación Censal de Estudiantes 2012 (ECE 2012). Lima: Ministerio de Educación del Perú. (Disponible en: http://www2.minedu.gob.pe/umc/ece2012/informes_ECE2012/Difusion/ECE%202012_Web_UMC.pdf)

MINEDU (2014). Resultados de la Evaluación Censal de Estudiantes 2014 (ECE 2014). Lima: Ministerio de Educación del Perú. (Disponible en: <http://umc.minedu.gob.pe/wp-content/uploads/2015/02/ECE-2014-Web-270215-27febv2.pdf>)

MINEDUC (2015). "Material Pedagógico". Ciudad de Guatemala: Ministerio de Educación de Guatemala-DIGEDUCA. Página Web. [Fecha de consulta: 14 Julio 2015]. (Disponible en: <http://www.mineduc.gob.gt/digeduca/>)

Naciones Unidas (2014): Open Working Group Proposal for Sustainable Development Goals, Agosto 2014. (Disponible en: www.undocs.org/A/68/970)

OCDE (2014). "Education at a Glance 2014: Brazil Country Note". Paris: OCDE. (Disponible en: <http://www.oecd.org/brazil/EAG2014-Country-Note-Brazil.pdf>)

OCDE (2012). Revisiones de la OCDE sobre la evaluación en la educación: México 2012. Paris: OECD Publishing.

OCDE (2011). Establecimiento de un marco para la evaluación e incentivos docentes: Consideraciones para México. Paris: OECD Publishing.

OEI (2012): Miradas sobre la Educación en Iberoamérica. Madrid: OEI/CEPAL/SEGIB. (Disponible en: <http://www.oei.es/miradas2012.pdf>)

SEP (2015). "Descarga serie histórica y pronósticos de la estadística del sistema educativo nacional". Ciudad de México: Secretaría de Educación Pública. (Disponible en: http://www.snie.sep.gob.mx/estadisticas_educativas.html)

Referencias, cont.

UNESCO Instituto de Estadísticas – UIS (2015). “Teachers by teaching level of education”. Página Web, Base de datos. [Fecha de consulta: 3 Agosto 2015].

Wolff, Laurence (2007): The Costs of Student Assessment in Latin America. Washington, DC: PREAL.

Zúniga Molina, Leonel y José Luis Gaviria (2010), Challenges and Opportunities for the Further Development of the ENLACE Assessment for Evaluation and Teacher Incentives in Mexico, un documento.

PÁGINAS WEB REVISADAS:

BRASIL

<http://www.inep.gov.br/>

COLOMBIA

<http://www.icfes.gov.co/>

<http://www.mineducacion.gov.co/>

GUATEMALA

<http://www.mineduc.gob.gt/>

MÉXICO

<http://www.inee.edu.mx/>

<http://www.sep.gob.mx/>

PERÚ

<http://www.minedu.gob.pe/umc>

URUGUAY

<http://www.anep.edu.uy/sea/>

Anexo 1: Lista de personas entrevistadas

Las siguientes personas fueron entrevistadas para este informe. Sus generosas contribuciones son aquí reconocidas. El informe presenta los hallazgos de esas entrevistas interpretados por los autores.

BRASIL

- María Helena Guimaraes de Castro. Ex Presidente del Instituto Nacional de Estudos e Pesquisas (INEP).
- Ruben Klein. Consultor para Fundação Cesgranrio y Presidente de la Asociación Brasileira de Evaluación Educativa, ABAVE (Associação Brasileira de Avaliação Educacional, ABAVE)

COLOMBIA

- Margarita Peña. Ex Presidente del Instituto Colombiano para el Fomento de la Educación Superior (ICFES).
- María Figueroa. Decano de la Facultad de Educación de la Universidad Externado de Colombia.

GUATEMALA

- Álvaro Fortín. Ex Director de la Dirección General de Evaluación e Investigación Educativa (DIGEDUCA).
- Verónica Spross. Investigadora Asociada del Centro de Investigaciones Económicas Nacionales (CIEN)

MÉXICO

- Margarita Zorrilla Fierro. Ex Directora General del Instituto Nacional para la Evaluación de la Educación (INEE).
- Marco Antonio Fernández. Director de Investigación de México Evalúa.

PERÚ

- José Rodríguez. Ex Director de la Oficina de Medición de la Calidad de los Aprendizajes (UMC).
- Patricia Arregui. Investigadora Principal del Grupo de Análisis para el Desarrollo (GRADE).

BANCO INTERAMERICANO DE DESARROLLO

- Emiliana Vegas. Jefe de la División de Educación.
- Alejandro Morduchowicz. Especialista Líder en Educación. Guatemala.

Anexo 2: Protocolo para las entrevistas (resumen de temas)

A. MARCOS INSTITUCIONALES

- Estabilidad
- Financiamiento y autonomía administrativa
- Recursos humanos
- Autonomía y capacidad para difundir los resultados
- Transparencia

B. INFORMES, DIFUSIÓN Y USO DE LOS RESULTADOS

- Coherencia entre los tipos de informes y usos previstos (el desarrollo del currículo, pedagogía, búsqueda de apoyo, formación docente, selección de los estudiantes, etc.)
- Adecuación de informes para diferentes audiencias (claridad, guía para la interpretación, la sensibilización, etc.)
- Información sobre los factores asociados y los modelos de valor agregado
- Entrega: plazos; alcance; regularidad
- Impacto: uso por parte de la escuela; diseño de políticas/programas; responsabilidad política

C. EVALUACIONES INTERNACIONALES

- Razones para la participación
- Desarrollo de capacidades técnicas
- Difusión de los resultados y el impacto en la opinión pública
- Uso específico de resultados

Anexo 3: Programas de evaluación 2009-2014

BRASIL

PAÍS Y AGENCIA EVALUADORA	AÑOS DE APLICACIÓN	GRADOS/AÑOS [1]	ÁREAS DISCIPLINARES [2]	TIPO DE ESTUDIO [3]	PRUEBAS INTERNACIONALES
Brasil Ministerio de Educación (MEC) INEP (Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira)	2009, 2011 (Aplicación bianual)	5, 9 S, 3 SS SAEB-ANEB	L,M	M	PISA 2009 2012 LLECE-TERCE 2013
	2009, 2011 (Aplicación bianual)	5, 9S SAEB-Prova Brasil	L,M	C	
	2013 (Aplicación bianual)	9S, 3 SS SAEB-ANEB	L,M,Cs	M	
	2013 (Aplicación bianual)	9S SAEB-Prova Brasil	L,M,Cs	C	
	2011, 2012, 2013, 2014	2S (en dos etapas: principios y fines de 2° año) Provinha Brasil	L	V	
	2009, 2010	2S (en dos etapas: principios y fines de 2° año) Provinha Brasil	L,M	V	
	2013, 2014 (Aplicación anual)	3 SAEB-ANA	L,M	C	
	2009-2014	Termino de la Educación Secundaria Superior ENEM	Cuatro testes y una redacción, comprendiendo diversos campos de conocimiento: L, M, LE, A, I, F, Q, B, G, H, FL, S	V	
	2009-2014	Jóvenes y Adultos Dos niveles: Ensino Fundamental (secundaria inferior) y Ensino Médio (secundaria superior) ENCCEJA	L, M, Cs, H, G	V	

COLOMBIA

PAÍS Y AGENCIA EVALUADORA	AÑOS DE APLICACIÓN	GRADOS/AÑOS [1]	ÁREAS DISCIPLINARES [2]	TIPO DE ESTUDIO [3]	PRUEBAS INTERNACIONALES
Colombia Ministerio de Educación ICFES (Instituto Colombiano para la Evaluación de la Educación)	2009	5,9 (SABER)	M,L,CN,	C y M [7]	PISA 2009 IEA-ICCS 2009
	2009-2014	11[3] (SABER 11: Examen de Estado de la Educación Media)	M, L, CS, CN, CC, LE, RC	C	
	2009-2014	Validación del Bachillerato [4]	L,M,B,F,Q,CS FL,LE	V	
	2009-2014	SABER PRO: [5] Examen de Estado de Calidad de la Educación Superior	Competencias genéricas necesarias para el adecuado desempeño profesional. Competencias específicas comunes a grupos de programas.	C (La presentación de SABER PRO es obligatoria como requisito de grado.)	
	2010-2014	Pre SABER 11[6]	M, L, CS, CN, CC, LE, RC	V	
	2011	5,9 (SABER)	L,M,CN,CC	M	PIRLS 2011
	2012	3,5,9 (SABER)	L,M,CN	C	PISA 2012
		5,9 (SABER)	CC	C	
	2013	3,5,9 (SABER)	L,M	s/d	LLECE-TERCE 2013
		5,9 (SABER)	CN	s/d	
	2014	3 (SABER)	L,M	s/d	
		5 (SABER)	L,M,CN	s/d	
		9 (SABER)	L,M,CN,EF	s/d	

GUATEMALA

PAÍS Y AGENCIA EVALUADORA	AÑOS DE APLICACIÓN	GRADOS/AÑOS [1]	ÁREAS DISCIPLINARES [2]	TIPO DE ESTUDIO [3]	PRUEBAS INTERNACIONALES
Guatemala Ministerio de Educación DIGEDUCA (Dirección General de Evaluación e Investigación Educativa)	2009	1,3,6 (Evaluación Nacional de Primaria)	L,M	M	IEA-ICCS 2009
		1 (Evaluación Muestral del Ciclo Básico)	L, M	M	
		3M (Evaluación de Tercero Básico)	L,M	C	
		3 (Evaluación Candidatos a Becas a la Excelencia)	L,M	V	
		Último año de secundaria divesificada (Prueba GRADUANDOS)	L,M	C	
	2010	1,3,6 (Evaluación Nacional de Primaria)	L,M	M	
		1, 2 (Evaluación Muestral del Ciclo Básico)	L, M	M	
		Último año de secundaria divesificada (Prueba GRADUANDOS)	L,M	C	
	2011*	1 (PAMI- Prueba de Habilidad de Matemática Inicial)	M	M	
		1 (LEE- Prueba de Habilidad Lecto-Escritura Emergente)	L	M	
		2 (ELGI- Evaluación de Lectura para Grados Iniciales)	L	M	
		1,2,3 (Evaluación Muestral del Ciclo Básico)	L, M	M	
		Último año de secundaria divesificada (Prueba GRADUANDOS)	L, M	C	
	2012	1,3,6 (Evaluación Nacional de Primaria)	L, M	M	
		1 (PAMI- Prueba de Habilidad de Matemática Inicial)	M	M	
		1 (LEE- Prueba de Habilidad Lecto-Escritura Emergente)	L	M	
		1,2,3 (Evaluación Muestral del Ciclo Básico)	L, M	M	
		3M (Evaluación de Escritura Tercero Básico)	W	M	
		Último año de secundaria divesificada (Prueba GRADUANDOS)	L, M	C	
		Último año de secundaria divesificada (Evaluación de Escritura a GRADUANDOS)	W	M	
	2013	1,3,6 (Evaluación Nacional de Primaria)	L, M	M	LLECE-TERCE 2013
3M (Evaluación de Tercero Básico)		L, M	C		
Último año de secundaria divesificada (Prueba GRADUANDOS)		L, M	C		
2014	3,6 (Evaluación Nacional de Primaria)	L, M	M		
	3M (Evaluación de Tercero Básico)	L, M	C		
	Último año de secundaria divesificada (Prueba GRADUANDOS)	L, M	C		

MÉXICO

PAÍS Y AGENCIA EVALUADORA	AÑOS DE APLICACIÓN	GRADOS/AÑOS [1]	ÁREAS DISCIPLINARES [2]	TIPO DE ESTUDIO [3]	PRUEBAS INTERNACIONALES	
México Secretaría de Educación Pública (SEP)	2009	3, 4, 5, 6 2S, 3S (Enlace Educación Básica)	L, M, FC y E	C		
		1S (Enlace Educación Básica)	L, M	C		
		Último grado de EMS (Enlace Educación Media Superior)	L, M	C		
	Prueba ENLACE (Evaluación Nacional del Logro Académico en Centros Escolares)	2010	3,4,5,6 2S, 3S (Enlace Educación Básica)	L, M, H		C
			1S (Enlace Educación Básica)	L, M		C
			Último grado de EMS (Enlace Educación Media Superior)	L, M		C
	2011	3,4,5,6 1S (Enlace Educación Básica)	L, M, G	C		
		2S, 3S (Enlace Educación Básica)	L, M	C		
		Último grado de EMS (Enlace Educación Media Superior)	L, M	C		
	2012	3,4,5,6 1S, 2S, 3S (Enlace Educación Básica)	L, M, Cs	C		
		Último grado de EMS (Enlace Educación Media Superior)	L, M	C		
	2013	3,4,5,6 2S, 3S (Enlace Educación Básica)	L, M, FCyE	C		
		1S (Enlace Educación Básica)	L, M	C		
		Último grado de EMS (Enlace Educación Media Superior)	L, M	C		
	2014	Último grado de EMS (Enlace Educación Media Superior)	L, M	C		
México INEE (Instituto Nacional para la Evaluación de la Educación)	2009 (Ciclo 2008-2009)	6 (EXCALE 06)	L, M, EC, CN	M	IEA-ICCS 2009 PISA 2009	
	2010 (Ciclo 2009-2010)	3 (EXCALE 03)	L, M, EC, CN	M		
		Último grado de EMS (EXCALE 12)	E, FC	M		
	2011 (Ciclo 2010-2011)	3 (Prescolar) (EXCALE 00)	L, M	M		
	Prueba EXCALE (Exámenes de la Calidad y el Logro Educativo)	2012 (Ciclo 2011-2012)	3S (EXCALE 09)	L, M, FCyE, Cs	M	PISA 2012
2013 (Ciclo 2012-2013)		6 (EXCALE 06)	L, M, FCyE, CN	M	LLECE-TERCE 2013	

PERÚ

PAÍS Y AGENCIA EVALUADORA	AÑOS DE APLICACIÓN	GRADOS/AÑOS	ÁREAS DISCIPLINARES [1]	TIPO DE ESTUDIO [2]	PRUEBAS INTERNACIONALES
Perú UMC (Oficina de Medición de la Calidad de los Aprendizajes)	2009	2 (ECE 2009)	L,M	C	PISA 2009
		4* (ECE 2009)	L2	C	
	2010	2 (ECE 2010)	L,M	C	
		4* (ECE 2010)	L2, L3	C	
	2011	2 (ECE 2011)	L,M	C	
		4* (ECE 2011)	L2	C	
	2012	2 (ECE 2012)	L,M	C	PISA 2012
		4* (ECE2012)	L2,L3	C	
	2013	2 (ECE 2013)	L,M	C	
		4* (ECE 2013)	L2	C	
	2014	2 (ECE 2014)	L,M	C	
		4* (ECE 2014)	L2, L3	C	

LEYENDAS**[1]**

M: Nivel Medio
S: Secundaria
SS: Secundaria Superior
EMS: Educación Media Superior

[2]

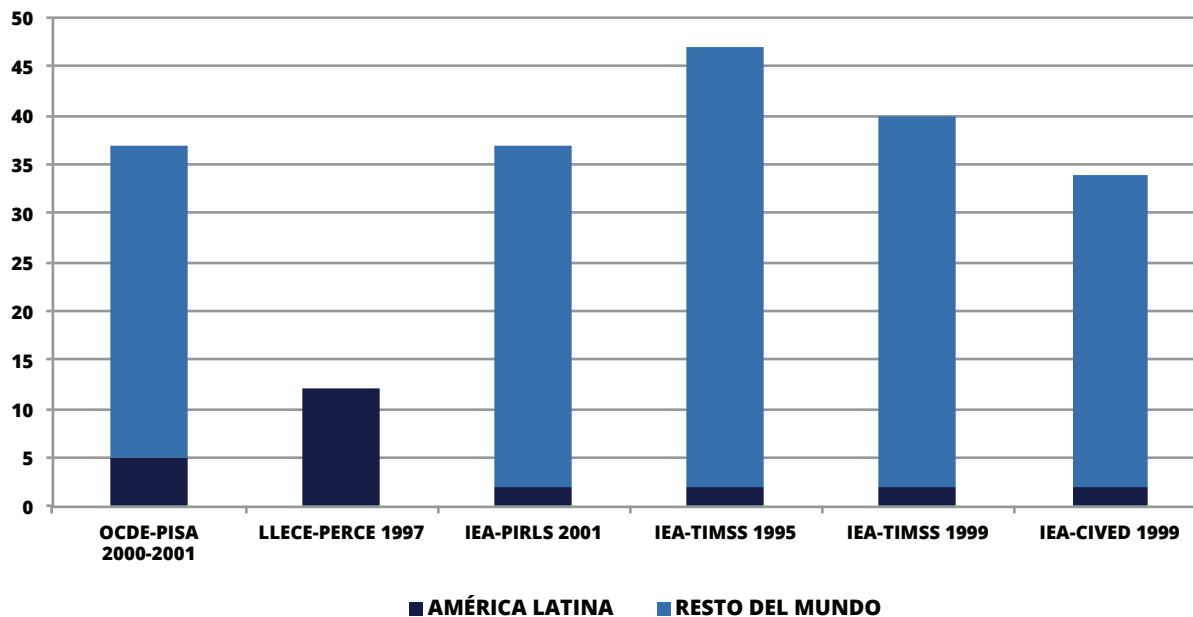
A: Arte
B: Biología
CC: Competencias Ciudadanas
Cs: Ciencias
CN: Ciencias Naturales
E: Expresión Escrita
EC: Educación Cívica
EF: Educación Financiera
F: Física
FC: Formación Ciudadana
FCyE: Formación Cívica y Ética
Fil: Filosofía
G: Geografía
H: Historia
I: Informática
L: Lenguaje (Español)
L2: Comprensión Lectora en Castellano como Segunda Lengua
L3: Comprensión Lectora en Cuatro Lenguas Originarias (Aimara, Awajún, Quechua Cusco-Collao, Shipibo Conibo)
LE: Lenguaje Extranjero
M: Matemática
Q: Química
RC: Razonamiento Cuantitativo
S: Sociología

[3]

C: Censal
M: Muestral
V: Voluntario

Anexo 4: Participación en pruebas internacionales

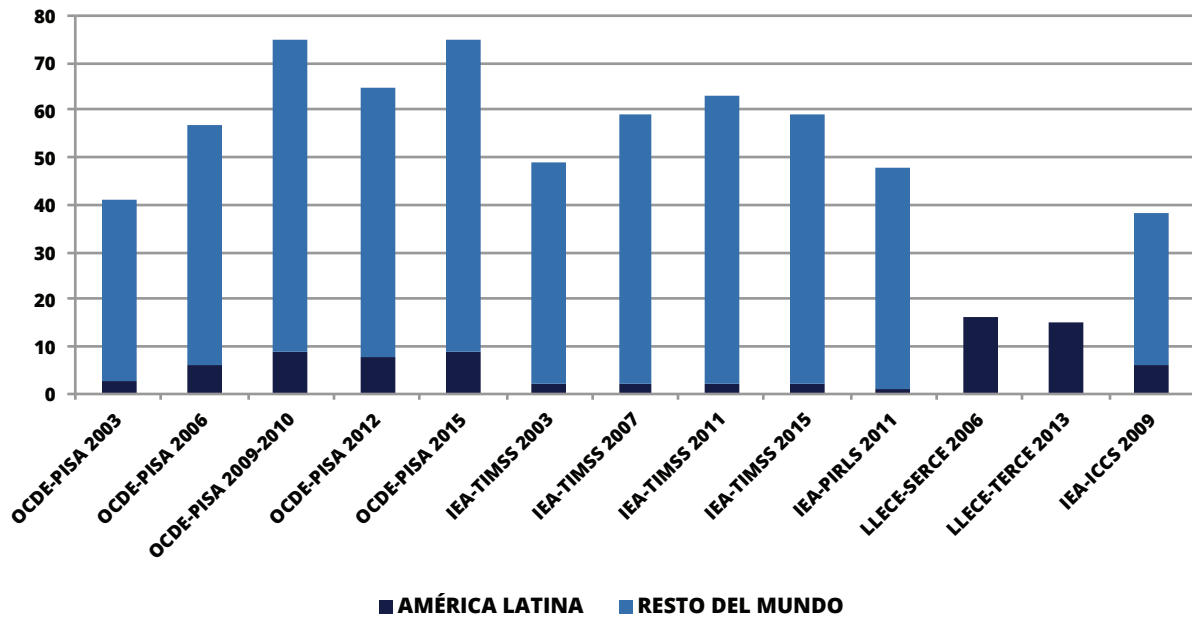
NÚMERO DE PAÍSES QUE PARTICIPARON EN ARREGUI Y FERRER (2003)



EVALUACIONES	AMÉRICA LATINA	RESTO DEL MUNDO
OCDE-PISA 2000-2001	5	32
LLECE-PERCE 1997	12	0
IEA-PIRLS 2001	2	35
IEA-TIMSS 1995	2	45
IEA-TIMSS 1999	2	38
IEA-CIVED 1999	2	32
Total de evaluaciones en la región: 6		

* Otros 11 países completaron la misma evaluación en el año 2002.

NÚMERO DE PAÍSES QUE PARTICIPARON EN LAS ÚLTIMAS APLICACIONES (2003-2015)



EVALUACIONES	AMÉRICA LATINA	RESTO DEL MUNDO
OCDE-PISA 2003	3	38
OCDE-PISA 2006	6	51
OCDE-PISA 2009-2010	9	66
OCDE-PISA 2012	8	57
OCDE-PISA 2015	9	66
IEA-TIMSS 2003	2	47
IEA-TIMSS 2007	2	57
IEA-TIMSS 2011	2	61
IEA-TIMSS 2015	2	57
IEA-PIRLS 2011	1	47
LLECE-SERCE 2006	16	0
LLECE-TERCE 2013	15	0
IEA-ICCS 2009	6	32
Total de evaluaciones en la región: 13		

TIMSS y PIRLS: No se incluyen entidades subnacionales.



Diálogo Interamericano

1211 Connecticut Ave. NW, Suite 510, Washington, DC 20036

Tel: 202-822-9002 / **Fax:** 202-822-9553

Correo electrónico: education@thedialogue.org

Web: thedialogue.org/agenda/programs/education