



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

La promesa de las evaluaciones de aprendizaje a gran escala

Reconocer los límites para generar oportunidades

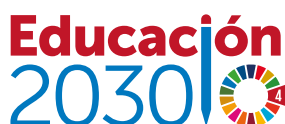
El Sector de Educación de la UNESCO

La educación es la prioridad principal de la UNESCO porque es un derecho humano esencial y es la base para consolidar la paz e impulsar el desarrollo sostenible. La UNESCO es la organización de las Naciones Unidas especializada en educación y su Sector de Educación proporciona un liderazgo mundial y regional en la materia, fortalece los sistemas educativos nacionales y responde a los desafíos mundiales de nuestra época mediante la educación, con un enfoque especial en la igualdad de género y África.



La Agenda Mundial de Educación 2030

En calidad de organización de las Naciones Unidas especializada en educación, la UNESCO ha recibido el encargo de dirigir y coordinar la Agenda de Educación 2030. Este programa forma parte de un movimiento mundial encaminado a erradicar la pobreza mediante la consecución, de aquí a 2030, de 17 Objetivos de Desarrollo Sostenible. La educación, fundamental para alcanzar todos estos objetivos, cuenta con su propio objetivo específico, el ODS 4, que se ha propuesto ***“garantizar una educación inclusiva, equitativa y de calidad y promover oportunidades de aprendizaje durante toda la vida para todos”***. El Marco de Acción de Educación 2030 ofrece orientación para la aplicación de este ambicioso objetivo y sus compromisos.



Publicado en 2019 por la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO), 7, place de Fontenoy, 75352 París 07 SP, Francia

© UNESCO 2019

ISBN 978-92-3-300120-6

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



La presente licencia se aplica exclusivamente al contenido del texto de la publicación. Para utilizar otros elementos cualesquiera (p. ej., imágenes, ilustraciones, gráficos) que no estén claramente identificados como propiedad de la UNESCO ni como material de dominio público, debe solicitarse una autorización previa a la UNESCO. (publication.copyright@unesco.org).

Título original: *The promise of large-scale learning assessments: Acknowledging limits to unlock opportunities.*

Publicado en 2019 por la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO).

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

Las ideas y opiniones expresadas en esta obra son las de los autores y no reflejan necesariamente el punto de vista de la UNESCO ni comprometen a la Organización.

Editado y corregido por Shereen Joseph

Fotografía de la portada: Getty / monkeybusinessimages

Diseñado e impreso por la UNESCO

Impreso en Francia

La promesa de las evaluaciones de aprendizaje a gran escala

Reconocer los límites para generar oportunidades

Prefacio

Por lo general, las pruebas estandarizadas se asocian a los exámenes escolares de fin de curso orientados a la certificación y selección de estudiantes para su avance a niveles educativos superiores. Más recientemente, los resultados de las pruebas estandarizadas se han utilizado para evaluar el funcionamiento de los sistemas educativos, como se refleja en la creciente relevancia de las evaluaciones de aprendizaje a gran escala (LSLA). Estas iniciativas, emprendidas a escala nacional o transnacional, no solo han llamado la atención sobre ciertos aspectos relacionados con los niveles de aprendizaje, sino también sobre los determinantes del aprendizaje, entre ellos la cualificación del profesorado, la calidad de los entornos escolares, el apoyo y la orientación parental, así como la salud social y emocional dentro y fuera de los centros educativos.

Nunca antes había habido tanta demanda de LSLA. Estas evaluaciones son hoy percibidas como fundamentales para mejorar la calidad educativa, y han contribuido en gran medida a fortalecer las políticas y estrategias orientadas a la mejora de un aprendizaje efectivo y pertinente. Siendo cada vez más usadas para asegurar la excelencia educativa, las LSLA están al centro de los debates nacionales, regionales y globales sobre la calidad y eficacia de los sistemas educativos. Pero nuestra búsqueda de la *excelencia* y la *calidad* educativas no puede realizarse a expensas de la *equidad*. La Agenda 2030 es clara al definir la equidad y la inclusión como perspectivas normativas centrales para el desarrollo educativo. Es un compromiso para asegurar que todos los niños, jóvenes y adultos tengan el derecho a un aprendizaje efectivo y pertinente. Si bien las LSLA son una herramienta central e indispensable para progresar hacia estas nuevas metas, en algunas circunstancias podrían socavar de manera accidental los compromisos de la agenda Educación 2030, centrados en garantizar un aprendizaje pertinente, igualitario y de calidad para todos.

La promesa de las evaluaciones de aprendizaje a gran escala: reconocer los límites para generar oportunidades aborda los aspectos más polémicos de estas evaluaciones estandarizadas. Sobre la base de la amplia experiencia de la UNESCO en esta esfera (desde su participación directa en la aplicación de las evaluaciones hasta su papel como transmisora de conocimientos y coordinadora de redes), esta publicación presenta la mirada crítica de la Organización respecto de estas iniciativas. Su objetivo es lograr un equilibrio en el debate sobre las LSLA examinando sus beneficios y concienciando al mismo tiempo sobre sus posibles riesgos y dificultades. Los debates de esta publicación se centrarán en las LSLA llevadas a cabo en el marco de la educación formal y en la etapa escolar.

Se espera que sean muchos los actores que puedan beneficiarse de las reflexiones contenidas en este documento. En particular, los gobiernos que efectúan LSLA o planean llevarlas a cabo; las organizaciones internacionales y regionales que prestan asistencia técnica para la planificación y la ejecución de dichas iniciativas; los académicos que se encargan de analizar críticamente estos

procesos; los gobiernos donantes, que cada vez demandan más este tipo de evaluaciones; y las empresas privadas a las que podría interesar introducirse en este nicho de mercado. Los debates planteados en esta publicación tienen por objetivo analizar y comprender las LSLA a partir de la exploración de una serie de cuestiones. ¿Por qué han adquirido importancia a lo largo del último decenio? ¿De qué manera podrían contribuir a cumplir los compromisos para 2030 de asegurar un aprendizaje inclusivo, igualitario y de calidad para todos? ¿Podrían perjudicar dichos compromisos? ¿Merece la pena invertir tiempo y recursos en ellas? ¿Están realmente dirigidas a todos los niños y sirven para promover la equidad? ¿Evalúan los conocimientos y las aptitudes pertinentes, o las competencias necesarias para desarrollarse en la sociedad actual?

El actual aumento de la participación en las LSLA se mantendrá en el futuro, no solo en cifras, sino también en sofisticación, a medida que aumenta su digitalización y adaptabilidad. Aunque esto suena prometedor, debemos estar atentos a los posibles riesgos que esto conlleva. ¿Cómo será el futuro de la evaluación y el uso de los datos resultantes? ¿Cómo garantizamos que los LSLA no continúen obstaculizando, de forma accidental, el progreso respecto de los compromisos de calidad para todos, o peor aún, exacerben las desigualdades sociales en términos de aprendizaje? Mirando al futuro, necesitamos asegurar una comprensión plena de las limitaciones de las LSLA (en relación con su diseño y posibles usos) y aplicar estos conocimientos al desarrollo de evaluaciones futuras.



Stefania Giannini,
Subdirectora General de Educación de la UNESCO

Agradecimientos

Esta publicación ha sido elaborada por el equipo de investigación y previsión en educación con contribuciones de colegas del Sector de Educación de la UNESCO, así como de institutos y oficinas regionales. La publicación fue coordinada por Maya Prince bajo las directrices de Sobhi Tawil.

Damos un agradecimiento especial a las expertas que han contribuido a las investigaciones y análisis documentales, a saber, Anne-Berit Kavli (Dirección de Educación y Capacitación de Noruega) y Clara Fontdevila (Universidad Autónoma de Barcelona).

Desde la UNESCO, queremos agradecer a Camilla Addey (Facultad de Magisterio de la Universidad de Columbia) y Seamus Hegarty (Universidad de Warwick), quienes se encargaron de las revisiones.

Gracias también a los pasantes de la UNESCO Zi Hu, Yutong Wang y Youngsun Cho, que ayudaron en la investigación de antecedentes.

Tabla de contenidos

Prefacio	3
Agradecimientos	5
Lista de cuadros, figuras y tablas	7
Siglas y acrónimos	8
Resumen	10
Introducción	14
1. El crecimiento de las evaluaciones de aprendizaje a gran escala	16
Contextualización del debate sobre las evaluaciones de aprendizaje a gran escala. . .	16
Un fenómeno en aumento	20
¿Qué impulsa el aumento de las evaluaciones de aprendizaje a gran escala?	23
2. La evaluación del aprendizaje en la agenda Educación 2030	28
Un mandato renovado para los datos sobre el aprendizaje	28
Contribuciones a la nueva agenda educativa global	32
3. Valorar más de lo que puede medirse: las limitaciones del diseño de las evaluaciones de aprendizaje a gran escala	36
Limitando la conceptualización de la educación y el aprendizaje	36
Limitando nuestra concepción de la calidad de la educación	39
Restringiendo los aspectos que valoramos	41
Restringiendo las personas que valoramos	49
4. Más allá de los datos como un fin en sí mismo: limitaciones en el uso de las evaluaciones de aprendizaje a gran escala	53
Usos de las evaluaciones de aprendizaje a gran escala: aclaración de términos y conceptos	53
Infrautilización de los datos de las evaluaciones	56
Sobreutilización de los datos de las evaluaciones	59
Asociación de las evaluaciones a la rendición de cuentas	65
Reflexiones finales	71
El cumplimiento de las promesas de las evaluaciones	71
El futuro de las evaluaciones	72
Referencias	74
Anexo 1: Evaluaciones transnacionales del aprendizaje	84
Anexo 2: Evaluaciones nacionales con marcos metodológicos comunes	88

Lista de cuadros, figuras y tablas

Cuadro 1	Beneficios percibidos asociados a la participación en las iniciativas transnacionales	24
Cuadro 2	Iniciativas recientes en el plano regional y global para promover las evaluaciones de aprendizaje a gran escala.	33
Cuadro 3	La Alianza Mundial para el Seguimiento del Aprendizaje	34
Cuadro 4	Coexistencia respetuosa: elemento de prueba del Estudio Internacional sobre Educación Cívica y Ciudadana de 2009	45
Cuadro 5	La adaptación de las pruebas: ¿un arma de doble filo?	51
Cuadro 6	Protegiendo a los alumnos, los docentes y las escuelas frente a las sanciones asociadas al desempeño	69
Figura 1	Los cuatro pilares de la educación	37
Figura 2	Porcentaje de alumnado por nivel de competencias lingüísticas y matemáticas alcanzado en los primeros cursos de la enseñanza primaria	64
Tabla 1	Sinopsis de las modalidades de evaluación	17
Tabla 2	El cambio hacia un enfoque centrado en el aprendizaje (publicaciones seleccionadas)	26
Tabla 3	Metas e indicadores del ODS 4 con relación al aprendizaje	29
Tabla 4	Información contextual reunida junto con los resultados del aprendizaje	31
Tabla 5	Ejemplos de iniciativas políticas en las que pueden influir las evaluaciones de aprendizaje a gran escala, por nivel	55

Siglas y acrónimos

ACER	Consejo Australiano de Investigación Educativa [<i>Australian Council of Educational Research</i>]
ADEA	Asociación pro Desarrollo de la Educación en África [<i>Association for the Development of Education in Africa</i>]
ANLAS	Análisis de los Sistemas Nacionales de Evaluación del Aprendizaje [<i>Analysis of National Learning Assessment Systems</i>]
ASER	Informe Anual sobre la Situación de la Educación [<i>Annual Status of Education Report</i>]
CONFEMEN	Conferencia de Ministros de Educación de los Países de Habla Francesa [<i>Conférence des ministres de l'Éducation des États et gouvernements de la Francophonie</i>]
DFAT	Departamento de Asuntos Exteriores y Comercios
DFID	Departamento para el Desarrollo Internacional del Reino Unido [<i>Department for International Development</i>]
EDS	Educación para el desarrollo sostenible
EGMA	Evaluaciones de matemáticas en los primeros cursos [<i>Early Grade Mathematics Assessment</i>]
EGRA	Evaluación de la lectura en los primeros cursos [<i>Early Grade Reading Assessment</i>]
ERCE	Estudio Regional Comparativo y Explicativo
GAML	Alianza Mundial para el Seguimiento del Aprendizaje [<i>Global Alliance to Monitor Learning</i>]
GEM	Informe de Seguimiento de la Educación en el Mundo [<i>Global Education Monitoring</i>]
ICCS	<i>International Civic and Citizenship Education Study</i> [Estudio Internacional sobre Educación Cívica y Ciudadana]
ICILS	<i>International Computer and Information Literacy Study</i> [Estudio Internacional sobre Competencia Digital y Tratamiento de la Información]
IEA	Asociación Internacional de Evaluación del Rendimiento Escolar [<i>International Association for the Evaluation of Educational Achievement</i>]
IEU	Instituto de Estadística de la UNESCO
IIPE	Instituto Internacional de Planeamiento de la Educación de la UNESCO
KIX	Intercambio de Conocimientos e Innovaciones [<i>Knowledge and Innovation Exchange</i>]
LAMP	Programa de Evaluación y Seguimiento de la Alfabetización [<i>Literacy Assessment and Monitoring Programme</i>]
LaNA	Evaluación de alfabetización y aritmética [<i>Literacy and Numeracy Assessment</i>]
LLECE	Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación
LSLA	<i>Large-scale learning assessments</i> [Evaluaciones de aprendizaje a gran escala]
NAEP	Evaluación Nacional de Progreso Educativo [<i>National Assessment of Educational Progress</i>]
NALA	Red para la Evaluación del Aprendizaje en África [<i>Network for African Learning Assessment</i>]
NEQMAP	Red para el Monitoreo de la Calidad Educativa en la Región Asia-Pacífico [<i>Network on Education Quality Monitoring in the Asia-Pacific</i>]

OAA	Optimización de las Evaluaciones para Todos
OCDE	Organización de Cooperación y Desarrollo Económicos
ODS	Objetivos de Desarrollo Sostenible
OIE	Oficina Internacional de Educación de la UNESCO
ONG	Organización no gubernamental
PAL	Red de Acción Ciudadana por el Aprendizaje [People's Action for Learning]
PASEC	Programa de Análisis de los Sistemas Educativos de la CONFEMEN [<i>Programme d'analyses des systèmes éducatifs de la CONFEMEN</i>]
PERCE	Primer Estudio Regional Comparativo y Explicativo
PIB	Producto interno bruto
PILNA	<i>Pacific Islands Literacy and Numeracy Assessment</i> [Evaluación de alfabetización y aritmética de las Islas del Pacífico]
PIRLS	<i>Progress in International Reading Literacy Study</i> [Estudio Internacional de Progreso en Comprensión Lectora]
PISA	<i>Programme for International Student Assessment</i> [Programa para la Evaluación Internacional de los Alumnos]
RTI	<i>Research Triangle Institute</i> [Instituto del Triángulo de Investigación]
SACMEQ	<i>Southern and Eastern Consortium for Monitoring Educational Quality</i> [Consortio de África Meridional y Oriental para la Supervisión de la Calidad de la Educación]
SEAMEO	<i>Southeast Asian Ministers of Education Organization</i> [Organización de Ministros de Educación de Asia Suroriental]
SEA-PLM	<i>Southeast Asia Primary Learning Metrics</i> [Medición del Aprendizaje en Primaria para el Sudeste Asiático]
SERCE	Segundo Estudio Regional Comparativo y Explicativo
SIMCE	Sistema de Medición de la Calidad de la Educación
TALENT	<i>Teaching and Learning: Educator's Network for Transformation</i> [Red de Docentes para la Transformación de la Enseñanza y el Aprendizaje]
TERCE	Tercer Estudio Regional Comparativo y Explicativo
TIC	Tecnología de la información y la comunicación
TIMSS	<i>Trends in International Mathematics and Science Study</i> [Estudio Internacional de Tendencias en Matemáticas y Ciencias]
UNESCO	Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura
UNICEF	Fondo de las Naciones Unidas para la Infancia
USAID	<i>United States Agency for International Development</i> [Agencia de los Estados Unidos para el Desarrollo Internacional]
VIH/sida	Virus de la inmunodeficiencia humana/síndrome de inmunodeficiencia adquirida

Resumen

Las evaluaciones de aprendizaje a gran escala (LSLA por sus siglas en inglés¹) han ido adquiriendo cada vez más importancia a lo largo de los últimos decenios. Constituyen una clase de prueba estandarizada a escala nacional o transnacional que ofrece una visión general acerca de los resultados de aprendizaje alcanzados por un grupo de estudiantes en un año concreto y en un determinado número de ámbitos de aprendizaje.

El uso de estas evaluaciones ha ido aumentando de manera constante (en la actualidad, se realizan en más de la mitad de los países de todo el mundo). Su alcance también ha aumentado. Además de la medición de las competencias de lectura y matemáticas, cada vez analizan un mayor número de ámbitos, como las aptitudes digitales, las competencias informáticas y de información, las habilidades socioemocionales o la comprensión de conceptos y cuestiones relacionadas con el civismo y la ciudadanía.

Las LSLA han ganado protagonismo en los debates en materia de educación celebrados en los planos local y mundial. Esta situación se ha visto reforzada gracias a una nueva perspectiva, inherente a la Agenda 2030 para el Desarrollo Sostenible, centrada en el aprendizaje igualitario, eficaz y pertinente para todos y al enfoque resultante relativo al uso de los datos para mejorar las políticas y las estrategias. Las LSLA están cada vez incorporadas a los programas de trabajo de organizaciones internacionales y regionales, y reciben el apoyo de organismos donantes a través de asistencia técnica y financiera.

Los beneficios y las ventajas de las LSLA han sido analizados de manera exhaustiva en la literatura especializada, y cada vez se otorga un mayor reconocimiento a la capacidad de los datos de las evaluaciones para guiar e informar la acción política de diversas formas. En consecuencia, a lo largo de los últimos decenios han crecido las expectativas sobre las LSLA como factores impulsores del cambio de las políticas. De hecho, se espera que tengan diversos usos, como el seguimiento, la rendición de cuentas, la elaboración de programas y los análisis. El potencial de dichas herramientas es mucho más amplio que la simple presentación de información. Así, pueden contribuir a entender qué esferas necesitan mejorar y ayudar a determinar cuáles son las intervenciones políticas más adecuadas, prometedoras y eficaces.

Sin embargo, tanto las evaluaciones nacionales como las transnacionales han suscitado inquietudes. Cada vez hay más evidencia que llama la atención sobre diversos efectos inesperados e incluso negativos derivados de dichos ejercicios. Sobre la base de un conjunto de evidencia diverso, dentro del que se incluye la literatura académica y la experiencia de numerosas organizaciones internacionales, asociados para el desarrollo y expertos en evaluaciones, en esta publicación se reflexiona sobre las posibles consecuencias imprevistas de las LSLA. Algunos problemas se derivan directamente de las características inherentes a su diseño, mientras que otros resultan de los usos (indebidos) de los datos para fundamentar la elaboración de programas y la formulación de políticas.

¹ *Large-scale learning assessments.*

Valorar más de lo que puede medirse: las limitaciones del diseño de las evaluaciones de aprendizaje a gran escala

Un primer conjunto de consecuencias imprevistas de las LSLA se deriva directamente de la manera en que están diseñadas. Cuatro de las principales dimensiones que es probable que se vean perjudicadas son nuestra conceptualización de la educación y el aprendizaje; nuestro entendimiento de la calidad de la educación; el rango de aptitudes, conocimientos y actitudes que valoramos; y nuestra atención a la inclusividad y el respeto de la diversidad.

Limitando la conceptualización de la educación y el aprendizaje

Se espera que la educación cumpla múltiples propósitos en el ámbito del desarrollo personal y social, tanto para los individuos como para las comunidades. Esta conceptualización de los propósitos múltiples e integrados de la educación se refleja en los cuatro pilares de la educación presentados por la Comisión Delors en 1996. Además, es probable que sea aún más importante hoy en día como medio para progresar hacia un desarrollo humano y social sostenible en un mundo en rápida evolución. Sin embargo, las LSLA suelen centrarse en un rango limitado de dimensiones de aprendizaje y abordar un número finito de los múltiples propósitos de la educación. Esto se debe en gran medida al diseño de las LSLA, que limita las aptitudes y las competencias mensurables a un pequeño número para posibilitar la comparación al nivel del sistema. Si bien esto no es inherentemente problemático, es probable que acabe limitando nuestro entendimiento de qué aspectos son valorables en la educación (sobre todo si las LSLA adquieren una importancia desproporcionada frente a otras formas de evaluación).

Limitando nuestra concepción de la calidad de la educación

Los resultados del aprendizaje medidos por las LSLA constituyen un indicador y dimensión clave de la calidad educativa. Sin embargo, los resultados del aprendizaje no pueden entenderse de manera aislada o como la única medición válida de la calidad de la educación. Más bien, deben concebirse en relación con otras dimensiones pertinentes, como los resultados a nivel social y los beneficios necesarios para la cohesión social, variables procedimentales relativas al proceso de enseñanza y aprendizaje, y los entornos del aula y la escuela. No obstante, los avances en estas dimensiones no se miden necesariamente mediante evaluaciones del aprendizaje. De hecho, todavía no está claro si este sería un cambio posible o incluso deseable. Si queremos capturar las distintas dimensiones de la calidad de la educación, las pruebas estandarizadas deben entenderse en combinación con una amplia gama de indicadores cualitativos y cuantitativos.

Restringiendo los aspectos que valoramos

El número limitado de ámbitos específicos que son objeto de medición puede considerarse una característica intrínseca de las LSLA. Esta restricción ha suscitado la preocupación de que dichos ejercicios no sirven para evaluar de manera adecuada el amplio número de esferas de competencia cultivadas a través de la educación. Incluso dentro de los ámbitos que sí se evalúan, las LSLA no prestan suficiente atención a todos los conocimientos, las aptitudes, las actitudes

y los valores que abarca la educación. Si bien algunas organizaciones están explorando nuevas estrategias para recoger las esferas de competencia menos frecuentemente exploradas, la viabilidad y la conveniencia de dichas mediciones sigue siendo cuestionable. Probablemente la atención limitada a algunas esferas de conocimientos podría solucionarse de una manera más adecuada a través de la combinación de las LSLA con otras formas de evaluación, de tal forma que los resultados del aprendizaje no sean en el único aspecto al que se presta atención.

Restringiendo las personas que valoramos

Las LSLA son herramientas altamente prometedoras como instrumentos para promover la equidad. Sin embargo, la exigencia de inclusividad no siempre se respeta en el diseño y la administración de las LSLA. Los niños, jóvenes y adultos en situación de vulnerabilidad y marginación pueden quedar excluidos durante el proceso de muestreo, pero también a nivel del aula o la escuela, como consecuencia del propio diseño o formulación de los elementos de las pruebas. Aunque cada vez se pone un mayor empeño en incluir a todos los estudiantes y registrar de manera adecuada su progreso, este aspecto sigue siendo un reto, pues es probable que la introducción de medidas especiales tenga como consecuencia incentivos perversos y problemas de comparabilidad.

Más allá de los datos como un fin en sí mismo: limitaciones en el uso de las evaluaciones de aprendizaje a gran escala

Las limitaciones asociadas a las LSLA también están relacionadas con el uso de sus resultados o la falta de ellos. Los datos de las LSLA cumplen varios propósitos, como el seguimiento de tendencias, la evaluación y el diseño de políticas, el establecimiento de agendas mediante la sensibilización acerca de determinados asuntos, y la responsabilización y rendición de cuentas de distintos actores.

A pesar de las buenas intenciones respecto del uso de los datos con el objetivo de mejorar los procesos educativos, surgen tres principales preocupaciones: su infrautilización, su sobreutilización, y su combinación con (o subordinación a) medidas de rendición de cuentas.

Infrautilización de los datos de las evaluaciones

Las LSLA solo pueden convertirse en un factor impulsor de cambios y mejoras si las partes interesadas en cuestión adoptan las medidas oportunas en función de sus resultados. Sin embargo, existen diversos factores que limitan el uso eficaz de los datos de las evaluaciones de tal manera que estos puedan servir para fundamentar de manera adecuada los debates sobre políticas y orientar la acción política. Entre ellos, se encuentran los limitados niveles de institucionalización de las prácticas de evaluación, así como el restringido sentido de apropiación respecto de la financiación, el diseño, la gestión de los datos y la difusión de las evaluaciones. Dado que se necesitan muchos recursos para la administración de las evaluaciones, el uso de los datos resultantes reviste una importancia primordial para que ese ejercicio no resulte inútil ni sea percibido como un derroche de recursos públicos.

Sobreutilización de los datos de las evaluaciones

Centrarse demasiado en los resultados de las evaluaciones u otorgarles un crédito excesivo puede inducir a errores y tener un efecto de desviación y distracción sobre las prioridades y los comportamientos de diversas partes interesadas. Esto ocurre en determinadas circunstancias y a través de diversos procesos. En primer lugar, las repercusiones de los datos de las evaluaciones en la formulación de las políticas se desaprovecharán si la acción política se orienta única o principalmente a mejorar las cifras en lugar de tratar de conseguir un cambio significativo. En segundo lugar, el uso de clasificaciones dentro de los países y entre ellos puede fomentar una dinámica de competición que, a su vez, promueva prácticas de préstamo de políticas mal fundamentadas, impulsadas por un deseo de emular a determinados países más que de aprender de ellos. En tercer lugar, los datos de las evaluaciones pueden fomentar un uso descuidado de la noción de causalidad que puede dar lugar a conclusiones mal fundamentadas sobre los efectos del aprendizaje. Por último, la incorporación de los datos de las evaluaciones en programas de financiación basados en los resultados, cada vez más utilizados por la comunidad de donantes, implica un riesgo importante, ya que podría dar lugar a estrategias orientadas exclusivamente a influir en los resultados.

Combinación de las evaluaciones con la rendición de cuentas

Muchos de los riesgos asociados a las LSLA son, de hecho, consecuencia de su uso en combinación con programas de rendición de cuentas de grandes repercusiones (*high-stakes*). Los efectos imprevistos de dichos planes se han analizado de manera exhaustiva en la literatura especializada. Algunas de las dificultades que se abordan con más frecuencia son la dinámica de la orientación de la enseñanza en función de las pruebas y la reducción de los planes de estudios. Sin duda, el entendimiento y el reconocimiento actuales de las consecuencias de las evaluaciones de menores repercusiones es limitado y merece una mayor investigación.

En general, las LSLA poseen un gran potencial, pero esconden algunos riesgos. Si bien dichas limitaciones no hacen que las evaluaciones sean inservibles ni inherentemente problemáticas, existen motivos para mantenerse vigilante en el diseño de las herramientas y el uso de la evidencia resultante. Como las LSLA son un elemento que se da cada vez más por sentado en la mayor parte de los sistemas educativos, es importante determinar qué esferas podrían beneficiarse de una mayor indagación, reflexión e investigación, como, por ejemplo, las cuestiones relacionadas con la transparencia y el sentido de la apropiación de los datos, el papel de otras modalidades de evaluación con fines de diagnóstico y el potencial de las variedades emergentes de LSLA.

Introducción

Las iniciativas de educación en el marco de la Agenda 2030 para el Desarrollo Sostenible otorgan una importancia renovada a la instauración de un aprendizaje igualitario, eficaz y pertinente para todos. Este énfasis implica que se debe poner un mayor empeño en realizar un seguimiento de los resultados de los procesos educativos, lo cual es coherente con el creciente reconocimiento del potencial de los datos de las evaluaciones para satisfacer diversos fines relacionados con las políticas públicas.

Las ventajas de las evaluaciones de aprendizaje a gran escala (LSLA²) se han consolidado con firmeza en un conjunto cada vez más amplio de literatura especializada. Los sistemas de las LSLA son un componente fundamental de seguimiento en los planos mundial, nacional e incluso local. La expansión experimentada por las LSLA en todo el mundo durante los dos últimos decenios es un indicador del importante papel que desempeñan. De hecho, han contribuido a mejorar las políticas y estrategias nacionales dirigidas a garantizar un aprendizaje eficaz y pertinente para todos. No obstante, al mismo tiempo que estas siguen satisfaciendo un conjunto cada vez mayor de propósitos, también surgen inquietudes en relación con sus posibles consecuencias imprevistas y el posible uso inadecuado de los datos procedentes de las evaluaciones a la hora de desarrollar políticas y prácticas educativas. Esta publicación contribuye al debate mundial examinando la evidencia existente en relación con los beneficios, los límites y los posibles efectos secundarios de las LSLA. En ella, se pide una mayor cautela a la hora de diseñar y aplicar las LSLA, así como al utilizar los datos resultantes. Debemos tomar estas precauciones si queremos cumplir nuestro compromiso colectivo de garantizar un aprendizaje eficaz y pertinente para todos.

Preguntas orientativas

En esta publicación se abordan las siguientes preguntas: ¿qué son las LSLA?; ¿cuáles son los factores impulsores de la expansión de dichas iniciativas?; ¿de qué forma se relacionan las LSLA con los compromisos mundiales para 2030?; y, lo que es más importante, ¿cuáles son las posibles consecuencias imprevistas del diseño de las LSLA y el uso de los datos resultantes para el desarrollo de políticas y prácticas? En términos generales, el objetivo de la publicación es lograr entender mejor cómo pueden mejorarse el diseño y el uso de las LSLA a fin de garantizar una educación equitativa y de calidad. En resumidas cuentas, lo que buscamos es comprender cómo desarrollar el potencial de las LSLA.

2 *Large-scale learning assessments.*

Organización

Capítulo 1

El capítulo presenta algunas aclaraciones conceptuales sobre la noción de las LSLA para contextualizarlas de manera adecuada dentro del alcance más amplio del seguimiento de la educación. Dicho capítulo contiene una sinopsis de la evolución de las LSLA a lo largo de los últimos decenios y describe la ampliación de su alcance y su aumento en número, importancia y visibilidad. Asimismo, se examinan las posibles lógicas de impulso detrás de esta expansión, así como la evolución de la cultura de evaluación y el cambio de enfoque en el debate mundial relativo a las políticas educativas – de uno centrado en el acceso a otro orientado en los logros de aprendizaje.

Capítulo 2

El capítulo sitúa las LSLA dentro del contexto de los principios y los compromisos que sustentan el Marco de Acción Educación 2030, así como dentro del debate educativo global sobre la cuestión del aprendizaje. En este capítulo se debate cómo la aprobación del Objetivo de Desarrollo Sostenible relativo a la educación (el ODS 4) ha impulsado una mayor atención a los resultados del aprendizaje y, además, se explora la relación que guardan las LSLA con el nuevo mandato y la manera en que lo apoyan. Por último, se examina el papel de los asociados para el desarrollo a la hora de garantizar que las evaluaciones reciban una atención adecuada como medio privilegiado para fomentar el progreso hacia la consecución del ODS 4.

Capítulo 3

El capítulo presenta el primer conjunto de consecuencias imprevistas asociadas al diseño y el desarrollo de las LSLA. En dicho capítulo se analizan cuatro dimensiones principales que corren el riesgo de verse perjudicadas como consecuencia del propio diseño de las LSLA. Estas dimensiones son nuestra conceptualización de la educación y el aprendizaje; nuestro entendimiento de la calidad de la educación; el rango de aptitudes, conocimientos y actitudes que valoramos; y nuestro compromiso con la inclusividad y el respeto de la diversidad.

Capítulo 4

El capítulo examina un segundo conjunto de consecuencias imprevistas asociadas a los usos de los datos de las LSLA por parte de diversas partes interesadas. Se abordan tres grupos principales de problemas. En primer lugar, se examina la infrautilización de los datos de las evaluaciones, en relación con lo cual se presta especial atención a las limitaciones que se derivan del restringido sentido de apropiación. En segundo lugar, se aborda la sobreutilización de los datos (es decir, la excesiva atención o el otorgamiento de un crédito excesivo a los resultados de las evaluaciones), que puede inducir a error o tener un efecto de distracción sobre las iniciativas políticas. Por último, se examina la combinación de la infrautilización o la sobreutilización de los datos con marcos de rendición de cuentas de grandes repercusiones (o la subordinación a estos).

En general, el objetivo de la publicación es poner de relieve las tensiones existentes entre el potencial de las LSLA como herramientas al servicio de la calidad, la equidad y la rendición de cuentas en la esfera de la educación, por una parte, y sus posibles consecuencias negativas en la práctica como consecuencia de un diseño o usos no adecuados, por la otra. En última instancia, para solucionar estas tensiones es necesario hacer un uso prudente y considerado de las LSLA con el fin de mitigar sus consecuencias potencialmente negativas.

1. El crecimiento de las evaluaciones de aprendizaje a gran escala

En este capítulo, se ofrece una sinopsis de la evolución de las evaluaciones de aprendizaje a gran escala (LSLA) a lo largo de los últimos decenios, en la cual se presta especial atención a su crecimiento en términos de alcance, número, importancia y visibilidad. En primer lugar, se presentan algunas aclaraciones conceptuales para contextualizar de manera adecuada las LSLA. A continuación, se examina la expansión y la evolución de las LSLA tanto a nivel tanto como transnacional. Por último, en este capítulo se reflexiona sobre los posibles factores que impulsan el crecimiento de la demanda de estas herramientas y el interés en ellas en los planos mundial y nacional.

Contextualización del debate sobre las evaluaciones de aprendizaje a gran escala

La importancia de las LSLA en los debates mundiales, nacionales e incluso locales relativos a la educación sólo ha aumentado durante los últimos decenios. Sin embargo, las LSLA deben entenderse como una variedad concreta de evaluaciones con características definidas que se integran en sistemas de evaluación más grandes y de gran complejidad. En este primer apartado, se ofrecen algunas aclaraciones conceptuales, mediante la definición de las LSLA y su contextualización en un marco más amplio.

Por motivos de simplicidad y claridad, este documento se centra en las LSLA realizadas con niños en edad escolar, aunque no necesariamente en un contexto escolar. La mayor parte de los debates que se abordan en esta publicación revisten especial importancia para la educación primaria y secundaria. No obstante, debe observarse que las LSLA que se centran en otros grupos de edad se ven afectadas por un conjunto de dificultades y limitaciones diferente que, si bien resulta interesante, escapa al alcance de este documento³.

¿Evaluación *del* aprendizaje o evaluación *para* el aprendizaje?

La evaluación es fundamental en cualquier proceso de aprendizaje. En muchas ocasiones, las personas expertas en pedagogía establecen una distinción entre la evaluación *para* el aprendizaje y la evaluación *del* aprendizaje. Dentro de la evaluación *para* el aprendizaje se engloban las evaluaciones formativas o en el aula que emplea el profesorado a fin de adaptar sus estrategias de enseñanza o como medio para asignar una calificación individual a los estudiantes al final de un determinado período de formación. Estas evaluaciones individuales están pensadas para realizar una valoración y un seguimiento de los conocimientos, las aptitudes, las actitudes

3 Este es el caso, por ejemplo, de los debates relativos a las mediciones de la alfabetización de los adultos (Addey, 2014) o sobre las implicaciones de una perspectiva de aprendizaje a lo largo de toda la vida sobre cuestiones de evaluación.

y los valores del alumnado en relación con el conjunto total de ámbitos académicos. Entre estas, se encuentran las evaluaciones de diagnóstico utilizadas para determinar qué alumnos tienen dificultades de aprendizaje y adaptar las actividades de formación y aprendizaje a sus necesidades (Clarke, 2012). Por otro lado, la evaluación *del* aprendizaje está relacionada con los exámenes o las evaluaciones sumativas que se utilizan para acreditar o seleccionar a alumnos de un curso determinado o una edad concreta a fin de que puedan pasar a un curso posterior, recibir formación superior o acceder a un puesto de trabajo. Son una forma de evaluación estandarizada que es coherente y está en armonía con el plan de estudios, aunque suele centrarse en asignaturas académicas troncales. En la **Tabla 1** se ofrece una sinopsis de estas distintas modalidades de evaluación. Debe observarse que la distinción entre la evaluación *para* el aprendizaje y la evaluación *del* aprendizaje (y la distinción equivalente de la evaluación formativa y la sumativas) es analítica y, como tal, no se corresponde exactamente con la práctica, si bien cada vez es mayor el número de evaluaciones híbridas en las que se combinan ambas funciones (Clarke, 2012).

Tabla 1. Sinopsis de las modalidades de evaluación

Evaluación interna o basada en la escuela	Evaluación formativa en el aula	
	Evaluación sumativa en el aula	
Evaluación externa o estandarizada	Exámenes principales (también conocidos como exámenes finales, públicos o de final de ciclo; son pruebas de grandes repercusiones para el alumnado)	
	Evaluaciones a gran escala o al nivel del sistema (suelen tener pocas repercusiones para el alumnado)	Evaluaciones transnacionales
		Evaluaciones nacionales y subnacionales

Celdas grises: Evaluación para el aprendizaje
Celdas blancas: Evaluación del aprendizaje

Fuente: Basado en Clarke, 2012; Fiske, 2000; OCDE, 2013.

Sistemas de evaluación del aprendizaje

Tanto la evaluación *del* aprendizaje como la evaluación *para* el aprendizaje son dimensiones fundamentales de los sistemas de evaluación del aprendizaje. Los sistemas de evaluación del aprendizaje consisten en un conjunto de políticas, estructuras, prácticas y herramientas para generar datos sobre los resultados del aprendizaje (Clarke, 2012; UNESCO, 2017a). Comprenden distintos tipos de evaluaciones que generan información en los planos individual, institucional o del sistema dirigida al profesorado, el alumnado, los perfiles encargados de la planificación y de la formulación de políticas, y otras partes interesadas. Estos sistemas pueden ayudar al alumnado a aprender, al profesorado a mejorar la enseñanza, a los cuadros encargados de la planificación a adoptar decisiones relativas a las asignaciones de recursos, y a los decisores políticos y los gobiernos a abordar las deficiencias del sistema o evaluar los programas educativos.

¿Qué son las evaluaciones de aprendizaje a gran escala?

Las LSLA son un subconjunto concreto de sistemas de evaluación del aprendizaje. Son evaluaciones a nivel del sistema que proporcionan un resumen de los resultados de aprendizaje de un grupo de alumnos determinado (basado en la edad o el curso), en un año académico

determinado y en un número limitado de ámbitos. Suelen clasificarse como evaluaciones nacionales o transnacionales (regionales/ internacionales)⁴. Las LSLA son uniformes y están estandarizadas en términos de contenido, proceso de administración, fechas y sistemas de puntuación (de hecho, suelen denominarse *pruebas estandarizadas*, sobre todo en los países y la literatura de tradición anglosajona). Por lo general, están basadas en muestreos, si bien un número cada vez mayor de países ha adoptado a lo largo de los últimos decenios un enfoque de base censal (Verger *et al.*, 2018). Pueden conducirse en la escuela o el hogar, pueden estar relacionadas con los planes de estudios o no, y puede tener consecuencias para el profesorado y las escuelas, pero tienen poco o ningún impacto sobre los candidatos que se presentan a las pruebas.

Las LSLA son evaluaciones a nivel del sistema que proporcionan un resumen de los resultados de aprendizaje de un grupo de alumnos determinado, en un año académico determinado y en un número limitado de ámbitos. Suelen clasificarse como evaluaciones nacionales o transnacionales.

En líneas generales, las LSLA son utilizadas por las autoridades de educación para determinar los niveles globales de resultados del alumnado. Ayudan a las autoridades educativas a realizar un seguimiento de los cambios en los resultados de aprendizaje a lo largo del tiempo y ponen de relieve las desigualdades en los resultados de aprendizaje entre los distintos grupos de población. Los datos de las LSLA, que determinan las correlaciones de los resultados de aprendizaje y ofrecen una visión más exhaustiva de las interacciones entre las diversas variables, permiten entender mejor las dinámicas que se encuentran detrás del desempeño de los sistemas educativos. De este modo, sirven de base para el diseño de políticas y estrategias dirigidas a mejorar los conocimientos y las competencias del estudiantado, así como la equidad en el aprendizaje.

Principios de rigor en el diseño y la aplicación

Una de las características que definen las LSLA (aunque no es exclusiva de estas evaluaciones) es el hecho de que se diseñan y se aplican de tal manera que cumplan ciertos criterios⁵. A pesar de la falta de normas internacionales examinadas acerca de las características que definen una LSLA adecuada (es decir, una evaluación que permita obtener datos fiables), sí existe un importante consenso entre las personas encargadas de desarrollar las pruebas, especialistas en estadística y expertas en psicometría sobre los requisitos técnicos de estas evaluaciones. Las LSLA deben desarrollarse y aplicarse sobre la base de tres principios, por lo menos.

En primer lugar, las LSLA deben ser sólidas desde el punto de vista técnico. Dicho en otras palabras, las metodologías de evaluación, el análisis y la interpretación de los datos deben seguir principios científicos. Por ejemplo, la evaluación debe incluir instrumentos contextuales correctamente formulados, herramientas de evaluación válidas y fiables, procesos de control de

4 Véase el Anexo 1 para consultar una lista de las evaluaciones transnacionales y otra información más detallada.

5 Aunque la estandarización del diseño y la administración de las evaluaciones constituye uno de los rasgos que mejor define las LSLA, debe observarse que dicha característica no es *exclusiva* de esta modalidad de evaluación. De hecho, la estandarización también es una característica clave de la mayor parte de los exámenes nacionales.

la calidad lingüística en relación con la traducción de los instrumentos, procedimientos rigurosos de muestreo científico y una selección de métodos analíticos adecuados (Consejo Australiano de Investigación Educativa (ACER), 2017)⁶.

En segundo lugar, en el marco de las LSLA, deben seguir un proceso de aplicación estandarizado. Por ejemplo, los datos deben recogerse en condiciones técnicas similares de una manera eficaz y fiable, con independencia del contexto de administración. Todo el personal que participa debe recibir una formación adecuada y deben administrarse los instrumentos pertinentes de conformidad con los procedimientos establecidos.

En tercer lugar, las LSLA deben diseñarse de tal manera que sean éticas y equitativas, y que incluyan a toda la población objetivo. Los principios éticos y de equidad deben aplicarse a todas y cada una de las fases de evaluación, entre las que se encuentran las siguientes:

- Fase de desarrollo (p. ej., debe hacerse un esfuerzo para reducir al mínimo el sesgo de medición⁷);
- Fase de aplicación (p. ej., debe tenerse en cuenta el bienestar de los participantes durante las evaluaciones, por ejemplo, a la hora de diseñar la extensión de la evaluación);
- Fases de análisis y difusión (p. ej., la confidencialidad de los participantes debe garantizarse en todo momento anonimizando los datos antes de analizarlos y hacerlos públicos).

Además, las evaluaciones deben ser inclusivas. Deben diseñarse de tal manera que sean pertinentes para el mayor número posible de miembros de la población a la que están dirigidas y de tal manera que ofrezcan una respuesta eficaz para todo el alumnado (UNESCO, 2017b). Así, la muestra debe reunir a un grupo representativo de alumnos con independencia de sus cualidades específicas, como sus orígenes (socioeconómicos u otros), su ubicación geográfica o su rendimiento académico. De igual modo, estos principios de equidad e inclusión tienen implicaciones, entre otras cosas, en relación con la traducción y la adaptación de los materiales de las pruebas para que se adecúen a grupos de población diferentes o para que incluyan a niños con necesidades especiales.

Las promesas de las LSLA

En la literatura especializada se han examinado extensamente los múltiples beneficios y ventajas de las LSLA. De hecho, hay un creciente reconocimiento del potencial de los datos de las evaluaciones para satisfacer diversos fines relacionados con las políticas públicas. Así pues, se ha observado que, si bien la evaluación forma parte de la fase de seguimiento y evaluación del ciclo de las políticas, sus usos van mucho más allá de esta etapa. Como explican Best et al. (2013), las LSLA también facilitan la evidencia clave necesaria para fundamentar el proceso de

6 En la actualidad, no existe ninguna norma internacional en la que se definan las características de una evaluación de aprendizaje “sólida desde el punto de vista técnico”. Sin embargo, en el documento *The Principles of Good Practice in Learning Assessment*, elaborado por el ACER en colaboración con el Instituto de Estadística de la UNESCO y la Alianza Mundial para el Seguimiento del Aprendizaje (ACER-IEU, 2017), se ofrece una articulación independiente de buenas prácticas que alberga la gran diversidad de actividades de LSLA llevadas a cabo en todo el mundo. Véase <http://uis.unesco.org/sites/default/files/documents/principles-good-practice-learning-assessments-2017-en.pdf> para obtener más información.

7 El sesgo de medición se produce cuando en una prueba o instrumento contextual se discrimina de manera sistemática a un grupo concreto por motivos no relacionados con el ámbito de aprendizaje evaluado ni los datos de antecedentes reunidos.

elaboración de los programas (es decir, la definición de las cuestiones prioritarias), orientar la formulación de las políticas y guiar su aplicación. De igual modo, Tobin *et al.* (2015) documentan cómo afectan los datos de las LSLA a diversas políticas educativas (en particular, políticas al nivel del sistema, entre ellas las reformas de los planes de estudios y el establecimiento de estándares de rendimiento, políticas de asignación de recursos y políticas de enseñanza y aprendizaje que afectan a las prácticas al nivel de escuela y aula)⁸. Como observan estos autores, las LSLA pueden utilizarse para una gran variedad de programas, como los de promoción de la calidad, la equidad y la rendición de cuentas. Asimismo, cada vez son más los organismos educativos especializados que ponen de relieve el potencial de las LSLA y defienden la continuidad de las inversiones en este tipo de instrumentos (p. ej., IEU, 2018a, o Banco Mundial, 2018).

Hay un creciente reconocimiento del potencial de los datos de las evaluaciones para satisfacer diversos fines relacionados con las políticas. Se ha observado que las LSLA facilitan la evidencia clave necesaria para fundamentar el proceso de elaboración de los programas y la definición de las cuestiones prioritarias, orientar la formulación de las políticas y guiar su aplicación.

En general, este conjunto de referencias bibliográficas deja claro que los beneficios de las LSLA son múltiples y que dichos instrumentos deben ser una característica fundamental de cualquier sistema educativo. La ausencia de datos sobre el aprendizaje implica diversos riesgos que comprometen el progreso hacia una mayor equidad y calidad. Sin embargo, las iniciativas dirigidas a definir y sintetizar las limitaciones, las dificultades y los riesgos asociados a las LSLA son más escasas y solo se emprenden en la fase de desarrollo. El objetivo del debate de esta publicación es, precisamente, profundizar en dichas limitaciones de una forma sistemática a fin de determinar cuáles son las condiciones ideales en las que es más probable que las LSLA estén a la altura de lo que prometen y repercutan en la calidad y la equidad de la educación.

Un fenómeno en aumento

En los dos últimos decenios, las LSLA, tanto nacionales como transnacionales, han ido adquiriendo cada vez más popularidad en todo el mundo y se han expandido mucho más allá de los países más ricos. De hecho, cada vez son más los países de ingresos medianos y bajos que invierten en la implementación de sistemas nacionales de evaluación y participan en evaluaciones transnacionales (UNESCO, 2015b). Entre 2013 y 2017, se administraron evaluaciones representativas a nivel nacional (entre ellas, evaluaciones transnacionales y nacionales) en al menos 177 países en el marco de la educación primaria y en al menos 150 países en el contexto del primer ciclo de la educación secundaria (IEU, 2019a). En este apartado se facilita una sinopsis de la expansión y la evolución de ambas variedades de evaluaciones.

⁸ De hecho, existen varias evaluaciones internacionales a gran escala que ofrecen análisis individuales o transnacionales sobre determinadas materias. Véase el anexo 1 de este informe para consultar una lista de enlaces a sitios web que contienen investigaciones consolidadas y análisis de evaluaciones transnacionales existentes.

Evaluaciones nacionales

A diferencia de los exámenes que se utilizan para acreditar los resultados de los estudiantes y seleccionarlos a fin de que puedan continuar con sus estudios, el principal propósito de las evaluaciones nacionales es ofrecer información a los encargados de la formulación de políticas en los planos nacional y sub-nacional acerca de los resultados de aprendizaje de una población concreta (Greaney, 2008; Clarke, 2012). Las evaluaciones nacionales suelen limitarse a unas pocas materias (Greaney, 2008), principalmente la lectura, las matemáticas y, en algunas ocasiones, las ciencias, las ciencias sociales y las lenguas extranjeras (Benavot y Köseleci, 2015)⁹. Estas evaluaciones se han extendido de 70 países entre 1990 y 1999 a 135 países en 2013 (Benavot y Köseleci, 2015), una expansión que se ha observado sobre todo en los países en desarrollo. Dentro de este crecimiento se engloban la aplicación de evaluaciones impulsadas desde la ciudadanía y las evaluaciones de la lectura en los primeros cursos (EGRA) durante el segundo y el tercer año de escolarización¹⁰, que se han administrado en más de 70 países en desarrollo en más de 100 idiomas. En una docena de países, aproximadamente, han pasado de ser pequeños cuestionarios modelo a evaluaciones representativas a nivel nacional (IEU, 2019a)¹¹. Si bien en el momento en que se escribió este informe no era posible ofrecer una cifra exacta del número de evaluaciones nacionales existentes, en la actualidad se están emprendiendo numerosas iniciativas para registrar los programas que se ejecutan a nivel nacional (hasta la fecha, la más exhaustiva de dichas iniciativas es la base de datos de evaluaciones del aprendizaje del IEU¹²).

Evaluaciones transnacionales

De igual modo, el número de países¹³ que participa en las evaluaciones transnacionales casi se ha duplicado desde sus inicios (1999 y 2000), y los últimos datos ponen de manifiesto una participación de 137 países en una o más evaluaciones transnacionales (IEU, 2018b)¹⁴. Por ejemplo, el número de países que participan en el Estudio Internacional de Tendencias en Matemáticas y Ciencias (TIMSS) pasó de 42 en 1995 a 64 en 2019 y, en el caso del Programa para la Evaluación Internacional de los Alumnos (PISA), el número de participantes ascendió de 42 en 2000 y 2001 a 79 en 2018; además, se añadieron otros 7 países a PISA para el Desarrollo (Raudonyte, 2019). También cabe mencionar que, si bien la mayor parte de los países participantes (el 60 %) siguen siendo de Europa y América del Norte (IEU, 2019a), la participación se ha expandido con el paso del tiempo a todas las regiones del mundo. Muchos países del Sur Global se están sumando en oleadas a los próximos ciclos de las evaluaciones transnacionales y se han desarrollado nuevas evaluaciones regionales. En África Subsahariana, por ejemplo, el

9 Algunas de las evaluaciones nacionales del aprendizaje más desarrolladas y consolidadas son la Evaluación Nacional de Progreso Educativo (NAEP) de los Estados Unidos de América y el Sistema de Medición de la Calidad de la Educación (SIMCE) de Chile.

10 Si bien las evaluaciones de la lectura en los primeros cursos son nacionales y a menudo sub-nacionales en términos de alcance (ya que son instrumentos específicos de cada país y no están concebidos para establecer comparaciones entre distintos países), el hecho de que se basen en un marco metodológico común ha llevado a algunos analistas a clasificarlas como evaluaciones internacionales o híbridas (véase Wagner, 2011; Raudonyte, 2019).

11 Para obtener más información sobre las evaluaciones de las iniciativas dirigidas por la ciudadanía y las evaluaciones de la lectura en los primeros cursos, consulte el anexo 2 del presente informe y el sitio web de Global Reading Network: <https://globalreadingnetwork.net/>

12 <http://uis.unesco.org/en/uis-learning-outcomes>

13 Debe observarse que, si bien las evaluaciones transnacionales suelen comparar distintos países, las regiones sub-nacionales también pueden constituir una unidad diferenciada de análisis, sobre todo en el caso de los sistemas federales o de gobiernos locales con altos niveles de autonomía administrativa.

14 Véase el anexo 1 para consultar una lista de las evaluaciones transnacionales y otra información más detallada.

Gabón, Guinea, Madagascar, Malí y la República Democrática del Congo se unirán al Programa de Análisis de los Sistemas Educativos (PASEC) de la Conferencia de Ministros de Educación de los Países de Habla Francesa (CONFEMEN) en 2019, mientras que, en América Latina, el Estado Plurinacional de Bolivia, Cuba y El Salvador se reincorporarán a la cuarta edición del Estudio Regional Comparativo y Explicativo (ERCE)¹⁵. En Asia y el Pacífico, la evaluación del programa Medición del Aprendizaje en Primaria para el Sudeste Asiático (SEA-PLM) se llevará a cabo por primera vez en 2019 en 11 países de la región.

Las evaluaciones del aprendizaje transnacionales, inicialmente diseñadas para evaluar las aptitudes básicas entre los niños de primaria y del primer ciclo de secundaria, se han ampliado para abarcar también a niños más pequeños, así como a la población joven y adulta, y analizar un abanico más amplio de ámbitos de aprendizaje. Entre ellas, se incluyen las competencias informáticas y de información, la comprensión de conceptos y cuestiones relacionados con la ciudadanía, y las actitudes y la acción cívicas.

Absteniéndose de participar, pero por poco tiempo

Aunque la participación en las evaluaciones de aprendizaje transnacionales va en aumento, es interesante observar que algunos países están absteniéndose de participar en ellas por motivos políticos, financieros o técnicos (Addey y Sellar, 2019)¹⁶. Por ejemplo, a mediados de la década del 2000, El Salvador dejó de participar en todas las evaluaciones internacionales, entre ellas el TIMSS, la evaluación del Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE) y el Programa de Evaluación y Seguimiento de la Alfabetización (LAMP), puesto que el nuevo gobierno electo interrumpió todas las actividades de evaluación al nivel del sistema. Argentina no participó en el PIRLS después de su primera edición en 2001. Estos cambios pueden haber surgido porque algunos países están más interesados en compararse con sus pares regionales, tal y como documentan Ferrer y Fiszbein (2015) respecto de los países de América Latina. Podría argumentarse que el PIRLS no presentaba ningún valor añadido, ya que la evaluación regional existente (LLECE) también mide las competencias lectoras de los niños en los primeros cursos de la educación primaria¹⁷. Estas tendencias sugieren que la participación en las evaluaciones transnacionales está motivada o impulsada por motivos tanto instrumentales como simbólicos.

Algunos países también han abolido sus evaluaciones nacionales del aprendizaje. Este es el caso de Sudáfrica, que llevaba a cabo una evaluación anual del aprendizaje dirigida a todo el estudiantado del primer al noveno curso. Esta evaluación estaba vinculada a la rendición de cuentas del profesorado y las escuelas, de manera que se adoptó la polémica decisión de

15 En el momento de la publicación del presente documento, aún es demasiado pronto para determinar qué países se unirán por primera vez al Estudio Internacional de Progreso en Comprensión Lectora (PIRLS) y al PISA de 2021.

16 En un plano diferente, también se han observado movimientos que llaman a abstenerse de participar en las evaluaciones en forma de iniciativas comunitarias dirigidas por la ciudadanía, sobre todo en los Estados Unidos de América. Estos movimientos pueden describirse como iniciativas colectivas que alientan a negarse a realizar pruebas estandarizadas. En general, han sido dirigidas por padres de niños en edad escolar (véanse Mitra *et al.*, 2016; o Pizmony-Levy y Saraisky, 2016). No obstante, este debate escapa al alcance de esta publicación.

17 Sin embargo, es importante observar que los países de América Latina están interesados en comparar la población que asiste a sus centros de secundaria con la de países altamente industrializados a través de evaluaciones como PISA, dirigida a jóvenes de 15 años (Ferrer y Fiszbein, 2015).

interrumpir la evaluación. Por otro lado, Uganda tuvo que cambiar su enfoque respecto a las evaluaciones anuales por problemas de financiación y optó por realizar estas pruebas cada tres años. Sin embargo, estos casos siguen siendo la excepción a la regla. A nivel mundial, las LSLA tienden a aumentar en número y a ampliarse en términos de alcance.

De hecho, parece probable que algunos países que han interrumpido las evaluaciones se reincorporen a los próximos ciclos de iniciativas transnacionales. Austria, Bosnia y Herzegovina, y Filipinas se reincorporarán al TIMSS de 2019 tras varios años sin participar en la evaluación. En América Latina, el Estado Plurinacional de Bolivia, Cuba y El Salvador se unirán al ERCE de 2019 después de haberse abstenido de participar en una o más ediciones de la evaluación. Argelia, que había dejado de tomar parte en evaluaciones transnacionales desde su participación en el TIMSS de 2011, se unió a PISA en 2015.

¿Qué impulsa el aumento de las evaluaciones de aprendizaje a gran escala?

Las razones que llevan a desarrollar una evaluación nacional del aprendizaje o a participar en iniciativas transnacionales son múltiples y están impulsadas por un amplio abanico de factores, a menudo relacionados entre ellos. Por tanto, no debe considerarse que la participación en las LSLA es un simple indicador de la riqueza o del estado de desarrollo ni que cumple una función particular dentro del ámbito de las relaciones internacionales. En este apartado, se reflexiona acerca de cuatro de los principales factores impulsores del aumento y la expansión de las LSLA: el creciente número de beneficios percibidos, la evolución a nivel mundial de la cultura de evaluación, el cambio en el enfoque de las políticas mundiales de educación, y las prioridades y demandas de los donantes para el desarrollo.

¿Cuáles son los beneficios percibidos?

Si bien algunas LSLA se realizan por mandato de donantes bilaterales y multilaterales (Benavot y Köseleci, 2015) o están impulsadas por preocupaciones relativas a los resultados y la rendición de cuentas (Kamens y McNeely, 2010), la participación también viene determinada por razones relacionadas con los beneficios percibidos del proceso de evaluación (véase el **Cuadro 1**). En particular, este es el caso de las iniciativas transnacionales. Entre estos beneficios se encuentran el desarrollo de relaciones internacionales positivas, la mejora de los resultados económicos y la fundamentación de planes de estudios y programas pedagógicos. Hay otros beneficios con un carácter más simbólico, que cuentan con un reconocimiento menos explícito, como la participación en un “ritual mundial de pertenencia”, la emulación de las denominadas “sociedades de referencia” y el empuje de las agendas personales, profesionales

Las razones que llevan a desarrollar una evaluación nacional del aprendizaje o a participar en iniciativas transnacionales son múltiples y están impulsadas por un amplio abanico de factores, a menudo relacionados entre ellos. No debe considerarse que la participación en las LSLA es un simple indicador de la riqueza o del estado de desarrollo ni que cumple una función particular dentro del ámbito de las relaciones internacionales.

o políticos de las personas encargadas de la formulación de políticas en el plano nacional (Addey, 2014; Addey *et al.*, 2017)¹⁸. Aunque existen diversas razones para participar en las evaluaciones, parece haber un denominador común: cada vez son menos los países que creen que es posible lograr el estado de una “buena sociedad” sin unos niveles educativos más elevados e iniciativas complementarias para el seguimiento de los resultados del aprendizaje (Kamens y McNeely, 2010).

Cuadro 1. Beneficios percibidos asociados a la participación en las iniciativas transnacionales

- **Facilitación del acceso a redes** en los planos regional e internacional, tanto para los equipos de investigación como para las personas encargadas de la formulación de políticas;
- **Fomento del diálogo** entre distintos grupos y facilitación de la aceleración del debate en materia de educación;
- **Generación de evidencia** sobre los fenómenos educativos y promoción del uso de indicadores técnicos y complejos sobre educación;
- **Promoción de reformas en las políticas reguladoras y de comportamiento**, entre otras cosas en materia de enseñanza, aprendizaje y evaluación;
- **Atracción de la atención de los medios de comunicación y aumento de la transparencia** en relación con los resultados de los sistemas educativos y el desarrollo del capital humano en contextos nacionales y transnacionales;
- **Desarrollo de las capacidades de los profesionales** que participan en la evaluación (desde un punto de vista técnico y operativo).

Fuente: Basado en Addey, 2014; Addey *et al.*, 2017; Díaz y Flores, 2008; Fenwick y Edwards, 2014; Lockheed *et al.*, 2015; Wagner *et al.*, 2012; OCDE, 2016b.

La evolución de la cultura de evaluación

El crecimiento de las LSLA también está asociado a un cambio de la cultura de evaluación. Con el paso del tiempo, las evaluaciones dirigidas por la Asociación Internacional de Evaluación del Rendimiento Escolar (IEA) y la OCDE han pasado de ser entendidos como estudios exhaustivos con finalidades de investigación y centrados en sistemas educativos concretos, a ser concebidos como iniciativas transnacionales orientadas hacia las políticas públicas y vinculadas a la rendición de cuentas y la competición (Pizmony-Levy, 2013; Addey, 2018). El cambio en la cultura de evaluación también se observa en otros aspectos. La evaluación del PASEC es un ejemplo en el que se pasó de un enfoque orientado a las evaluaciones nacionales individuales a otro centrado en los estudios comparativos. Entre 1991 y 2012, el PASEC llevó a cabo evaluaciones nacionales individuales en 24 países y territorios de habla francesa en África, Oriente Medio y Asia Oriental. Hasta entonces, los datos no eran comparables entre los distintos países. No fue hasta 2014 que el PASEC cambió su enfoque para llevar a cabo su primera evaluación transnacional en

¹⁸ En este sentido, el aumento de las LSLA no puede disociarse de otros cambios más amplios y, en particular, del surgimiento de nuevos paradigmas hegemónicos en materia de políticas públicas, como la nueva gestión pública (Addey *et al.*, 2017). Los modelos de políticas establecidos por este enfoque al servicio público ponen de relieve la función evaluativa del Estado (en materia de educación u otras esferas). Así, en un contexto en que los países se enfrentan a unos niveles de presión cada vez mayores para reformar sus sistemas educativos de acuerdo a estos principios, la adopción de las LSLA se considera una manera relativamente económica y (políticamente) ventajosa de lograr este objetivo (Verger *et al.*, 2018).

diez países. Esta evolución de las culturas de evaluación es consecuencia de la intensificación de la globalización y refleja el creciente interés en los mandatos mundiales (Greaney, 2008) y el creciente énfasis en las LSLA en el discurso educativo global.

Un cambio de foco en el discurso educativo global – del *acceso al aprendizaje*

El debate mundial en materia de educación, sobre todo en los años previos a la aprobación de la Agenda 2030, pasó de un enfoque centrado en el *acceso a la escolarización* a otro orientado al *aprendizaje*. En efecto, se ha observado un cambio general en el debate global en materia educativa, que pasó de prestar atención a los *insumos* de la educación, como las tasas de escolarización del alumnado, las instalaciones escolares, los materiales de las aulas, y la educación y la formación del profesorado, a interesarse por los *resultados* de la educación, es decir, los conocimientos, las aptitudes, las actitudes y los valores que el alumnado adquiere como *consecuencia* de la educación. Este cambio puede observarse en documentos de estrategia elaborados recientemente por asociados internacionales para el desarrollo (véase la **Tabla 2**), los cuales, junto con otros agentes, defienden un cambio de paradigma para pasar de garantizar un *acceso equitativo* a asegurar un *aprendizaje pertinente y eficaz para todos* (UNESCO-OIE, 2013; Barrett, 2016). La *equidad* y la *calidad* se han convertido en las protagonistas de los debates mundiales relativos a la educación, y las LSLA han pasado a considerarse herramientas para ayudar a los países a alcanzar su objetivo de lograr calidad y equidad en la educación¹⁹.

La equidad y la calidad se han convertido en las protagonistas de los debates mundiales relativos a la educación, y las LSLA han pasado a considerarse herramientas para ayudar a los países a alcanzar su objetivo de lograr calidad y equidad en la educación.

Las negociaciones en torno a la agenda Educación 2030 desempeñaron un papel crucial a la hora de garantizar y consolidar la interrelación de la calidad de la educación, los resultados del aprendizaje y las LSLA. Esto se debió, sobre todo, a la selección de indicadores para las distintas metas. De hecho, en varios informes y declaraciones publicados durante el período previo a la aprobación de la nueva agenda se hizo hincapié en una conceptualización distinta de la calidad. Si bien en algunos de ellos se eligió un restringido respecto de la calidad, que la equiparaba con los resultados en materia de aritmética y alfabetización, la interpretación que finalmente prevaleció fue significativamente amplia y exhaustiva en términos de alcance. No obstante, la traducción de la lista final de metas en indicadores implicó una reducción relativa de las ambiciones originales, al insistirse en la medición de las competencias en lectura y matemáticas como un indicador adecuado (King, 2017; Unterhalter, 2019).

¹⁹ Aunque las LSLA han ocupado un lugar central en los debates internacionales, debe observarse que la consideración y la conceptualización de estas herramientas como medio clave para el seguimiento y fomento del progreso educativo se remontan, por lo menos, a los inicios del programa Educación para Todos. En el informe *Evaluación del aprovechamiento escolar*, publicado en el 2000 en el marco de la serie *Situación y tendencias* (Fiske, 2000), ya se indagaba sobre el papel fundamental de la evaluación de los resultados del alumnado como herramienta para promover una educación de calidad para todos. En este sentido, cabe observar que en dicho informe ya se desarrolla y se analiza la mayor parte de las cuestiones en torno a las cuales siguen girando los debates en la actualidad.

Tabla 2. El cambio hacia un enfoque centrado en el aprendizaje (publicaciones seleccionadas)

Publicación/Iniciativa	Organización	Interés principal
<i>Aprendizaje para todos: invertir en los conocimientos y las capacidades de las personas para fomentar el desarrollo. Estrategia de Educación 2020 del Banco Mundial (2011)</i>	Grupo Banco Mundial	Enfoque centrado en garantizar que todos los niños y jóvenes, además de asistir a la escuela, adquieran los conocimientos y las aptitudes que necesitan para llevar vidas saludables y productivas, y conseguir un empleo de calidad.
<i>LEARNING to Realize Education's Promise. Informe sobre el desarrollo mundial de 2018 (2018)</i>	Grupo Banco Mundial	Resultados del aprendizaje.
Iniciativa Mundial La Educación ante Todo (2012-2016)	Iniciativa Mundial de Educación del Secretario General de las Naciones Unidas	Escolarización de todos los niños, calidad del aprendizaje, ciudadanía mundial.
Comisión Especial sobre Métricas de los Aprendizajes. Acceso más aprendizaje (2012-2016)	Instituto de Estadística de la UNESCO y Centro de Educación Universal de Brookings	Fortalecimiento de los sistemas de evaluación y del uso de los datos de las evaluaciones.
<i>La generación del aprendizaje. Invertir en educación para un mundo en proceso de cambio (2016)</i>	Comisión de Educación (Comisión Internacional sobre la Financiación de las Oportunidades de Educación Mundial)	Mobilización y despliegue de recursos para garantizar que todos los niños y jóvenes tengan la oportunidad de recibir educación y adquirir las aptitudes que necesitarán para convertirse en adultos productivos y exitosos.

Nota: Esta selección de iniciativas no es exhaustiva. Está pensada para ejemplificar el enfoque centrado en el *aprendizaje* adoptado por distintas iniciativas, mecanismos de financiación y estrategias organizativas.

El énfasis en la producción y el uso de datos

En la actualidad, los asociados para el desarrollo y los organismos donantes multilaterales establecen cada vez con mayor frecuencia algún tipo de evaluación del aprendizaje para sus préstamos y ayudas (Kamens y McNeely, 2010). Es habitual que los organismos donantes utilicen los resultados de las LSLA para evaluar la eficacia de las políticas y prácticas educativas. Los datos de las evaluaciones pueden ayudar a llevar a cabo un seguimiento de la evolución de las lagunas de aprendizaje que existen entre los distintos grupos y dentro de estos a partir de factores como el género, el nivel ingresos, la residencia y la lengua materna. El Plan Estratégico para 2020 de la Alianza Mundial para la Educación, por ejemplo, incorporó el uso de financiación basada en resultados con el fin de ayudar a los gobiernos a crear sistemas de evaluación sólidos y los anima a efectuar todo tipo de evaluaciones de aprendizaje para realizar un seguimiento de los resultados del aprendizaje (Alianza Mundial para la Educación, 2016). Para acceder a la financiación de la Alianza Mundial para la Educación, todos los beneficiarios deben demostrar que poseen sistemas de datos de buena calidad o una estrategia para desarrollarlos (Alianza Mundial para la Educación, 2018). De manera similar, la estrategia educativa Aprendizaje para Todos del Banco Mundial presta apoyo a las iniciativas gubernamentales a fin de aumentar los datos disponibles en materia de aprendizaje y aptitudes, y de desarrollar una cultura de seguimiento y evaluación de los resultados (Banco Mundial, 2011). El Marco de Aprendizaje

2030 de la OCDE (OCDE, 2018a) y la Estrategia de Educación de la UNESCO (2014-2021) incluyen también un importante énfasis en la evaluación del aprendizaje como forma de realizar un seguimiento de las aptitudes básicas y transferibles, así como en la utilización de datos para tomar decisiones fundamentadas que mejoren la enseñanza y el aprendizaje.

Se observan prioridades similares en las estrategias de desarrollo de algunos de los principales donantes bilaterales en el ámbito educativo. Por ejemplo, en su nueva Estrategia de Educación, el Departamento para el Desarrollo Internacional del Reino Unido (DFID) subraya su compromiso de trabajar con los encargados de la toma de decisiones para mejorar la calidad y el uso de los datos sobre educación, al otorgar prioridad a las mediciones adecuadas del aprendizaje a nivel nacional y transnacional (DFID, 2018). La Estrategia de Educación (2015-2020) del Departamento de Relaciones Exteriores y Comercio de Australia también incluye un importante enfoque centrado en la mejora de los sistemas de evaluación del aprendizaje y el uso de evidencia para fundamentar las buenas políticas y prácticas (Departamento de Relaciones Exteriores y Comercio de Australia, 2015). Asimismo, los gobiernos del Canadá, los Estados Unidos de América, Finlandia, Noruega y Suecia recalcan la importancia de las evaluaciones del aprendizaje como herramientas esenciales a la hora de realizar un seguimiento del progreso en el aprendizaje (Organismo Noruego de Cooperación para el Desarrollo (NORAD), 2018; Agencia Sueca de Cooperación Internacional para el Desarrollo (ASDI), 2018; Departamento de Relaciones Exteriores y Comercio del Canadá, 2014; Agencia de los Estados Unidos para el Desarrollo Internacional (USAID), 2011; Ministerio de Educación de Finlandia, 2006).

2. La evaluación del aprendizaje en la agenda Educación 2030

Si bien las evaluaciones de aprendizaje a gran escala (LSLA) y los resultados del aprendizaje han cobrado impulso recientemente en la comunidad para el desarrollo, la adopción del ODS 4 ha impulsado una atención sin precedentes a estos componentes educativos. En este capítulo, se reflexiona sobre la relación entre las LSLA y la agenda Educación 2030, y sobre cómo diversos asociados para el desarrollo promueven un mayor énfasis global en los datos sobre el aprendizaje como medio de fomentar el progreso hacia la consecución del ODS 4.

Un mandato renovado para los datos sobre el aprendizaje

La agenda Educación 2030 ha sido esencial para consolidar la atención prestada a las evaluaciones y los resultados del aprendizaje, así como su carácter central, lo que ayuda a concebir las LSLA como un factor indispensable para realizar un seguimiento del aprendizaje y del avance hacia las metas nuevas. En el presente apartado, se aborda el carácter fundamental de las LSLA a la hora de promover esta nueva visión de la educación, no solo como herramienta para el seguimiento y la elaboración de informes, sino también para orientar y fomentar las iniciativas normativas y, en última instancia, posibilitar el logro del ODS 4.

La visión de la agenda Educación 2030

¿Qué relación tienen las LSLA con la agenda Educación 2030? Como se indicó en la Declaración de Incheon de 2015, la agenda Educación 2030 se inspira en una visión humanística de la educación y el desarrollo basada en los principios del respeto por los derechos humanos y la dignidad humana; la justicia social; la inclusión; la protección; la diversidad cultural, lingüística y ética; y la rendición de cuentas y la responsabilidad compartidas. Reafirma el principio de la educación como un bien público, un derecho humano fundamental y la base para garantizar el disfrute de otros derechos. Por otra parte, otorga a la educación una posición esencial para la realización humana y el desarrollo sostenible en un marco donde el crecimiento económico se rige por una preocupación esencial por la justicia social y la gestión ambiental. La educación en la Agenda 2030 para el Desarrollo Sostenible ha redirigido la atención hacia la equidad y la inclusión, además de hacia la calidad y la relevancia de la educación como parte de un enfoque de aprendizaje a lo largo de toda la vida. Esto se expresa fundamentalmente en el Objetivo de Desarrollo Sostenible 4 (ODS 4), que trata de “garantizar una educación inclusiva, equitativa y de calidad y promover oportunidades de aprendizaje durante toda la vida para todos” (UNESCO, 2016c, pág. 7).

La agenda Educación 2030 ha sido esencial para consolidar la atención prestada a las evaluaciones y los resultados del aprendizaje, así como su carácter central, lo que ayuda a concebir las LSLA como un factor indispensable para realizar un seguimiento del aprendizaje y del avance hacia las metas nuevas.

El énfasis del ODS 4 en los resultados y productos de la educación

El ODS 4 se centra en los resultados y productos de los procesos educativos. Hace hincapié en la adquisición eficaz de conocimientos relevantes, competencias y aptitudes básicas y transferibles como fundamento para el aprendizaje permanente, y también en la relevancia del aprendizaje para el ámbito del trabajo y para la vida cívica, social y cultural (UNESCO, 2016a). Esto se refleja claramente en las metas e indicadores mundiales basados en resultados que se propusieron para el seguimiento de los avances hacia el logro del ODS 4 (véase la **Tabla 3**).

En vista del carácter fundamental de los resultados del aprendizaje en la nueva agenda, los datos sobre el aprendizaje son cada vez más percibidos como un componente esencial de los mecanismos de seguimiento destinados a medir el avance hacia la consecución del ODS 4, en el plano mundial, regional y nacional.

Tabla 3. Metas e indicadores del ODS 4 con relación al aprendizaje

Meta 4.1. De aquí a 2030, asegurar que todas las niñas y todos los niños terminen la enseñanza primaria y secundaria, que ha de ser gratuita, equitativa y de calidad y producir resultados de aprendizaje pertinentes y efectivos.	
Indicador 4.1.1	Porcentaje de niños y jóvenes: a) en los grados 2/3; b) al final de la enseñanza primaria; y c) al final del primer ciclo de la enseñanza secundaria, que han alcanzado al menos un nivel mínimo de competencia en i) lectura y ii) matemáticas, por sexo.
Meta 4.2. De aquí a 2030, asegurar que todas las niñas y todos los niños tengan acceso a servicios de atención y desarrollo en la primera infancia y educación preescolar de calidad, a fin de que estén preparados para la enseñanza primaria.	
Indicador 4.2.1	Porcentaje de niños menores de 5 años cuyo desarrollo se encuentra bien encauzado en cuanto a la salud, el aprendizaje y el bienestar psicosocial, por sexo.
Meta 4.4. De aquí a 2030, aumentar considerablemente el número de jóvenes y adultos que tienen las competencias necesarias, en particular técnicas y profesionales, para acceder al empleo, el trabajo decente y el emprendimiento.	
Indicadores 4.4.1	Porcentaje de jóvenes y adultos con conocimientos de tecnología de la información y la comunicación (TIC), por tipo de conocimiento técnico.
4.4.2	Porcentaje de jóvenes y adultos que han alcanzado al menos un nivel mínimo de competencia en alfabetización digital.
Meta 4.6. De aquí a 2030, asegurar que todos los jóvenes y una proporción considerable de los adultos, tanto hombres como mujeres, estén alfabetizados y tengan nociones elementales de aritmética.	
Indicador 4.6.1	Porcentaje de población en un grupo de edad determinado que alcanza por lo menos un nivel fijo de competencia funcional en a) alfabetización y b) aritmética, por sexo.
Meta 4.7. De aquí a 2030, asegurar que todos los alumnos adquieran los conocimientos teóricos y prácticos necesarios para promover el desarrollo sostenible, entre otras cosas mediante la educación para el desarrollo sostenible y los estilos de vida sostenibles, los derechos humanos, la igualdad de género, la promoción de una cultura de paz y no violencia, la ciudadanía mundial y la valoración de la diversidad cultural y la contribución de la cultura al desarrollo sostenible.	
Indicadores 4.7.4	Porcentaje de estudiantes por grupo de edad (o nivel educativo) que muestran una comprensión adecuada de las cuestiones relacionadas con la ciudadanía mundial y la sostenibilidad.
4.7.5	Porcentaje de estudiantes de 15 años que muestran dominio del conocimiento de geociencias y ciencias ambientales.

Fuente: UNESCO 2016a.

Nota: Es posible consultar la lista integral de las diez metas del ODS 4 y los 43 indicadores correspondientes en Desglosar el Objetivo de Desarrollo Sostenible 4²⁰. Los datos sobre los 43 indicadores temáticos están disponibles en la base de datos del IEU. <http://data.uis.unesco.org/>

20 Véase *Desglosar el Objetivo de Desarrollo Sostenible 4*: https://unesdoc.unesco.org/ark:/48223/pf0000246300_spa.

Una herramienta indispensable para el seguimiento de los indicadores del ODS 4

Dado que algunas de las metas y los indicadores mundiales del ODS 4 se basan en resultados, las LSLA han pasado a una forma relevante de aportar evidencia oportuna y fiable para el seguimiento de los compromisos relacionados con el aprendizaje (IEU, 2018c). De hecho, se espera que las distintas metas vinculadas a resultados que se recogen en la tabla 3 sean objeto de un seguimiento por medio de algún tipo de evaluación a nivel del sistema de los logros de aprendizaje. Por consiguiente, se considera que las LSLA son herramientas indispensables para llevar a cabo un seguimiento del aprendizaje, al menos hasta que se desarrollen nuevas metodologías que integren otras fuentes de datos (p. ej., datos administrativos, exámenes nacionales, censos o encuestas de hogares) y estimaciones o proyecciones.

Más allá de las metas educativas globales

Las mediciones del aprendizaje y las LSLA son fundamentales para la agenda Educación 2030, ya que permiten elaborar informes sobre una amplia variedad de metas relacionadas con resultados, pero su potencial va mucho más allá de este propósito²¹. Las mediciones del aprendizaje también proporcionan evidencia esencial sobre los determinantes de los logros académicos de los alumnos, así como sobre los efectos y la eficiencia de las iniciativas políticas destinadas a respaldar y mejorar dichos logros.

Las LSLA se pueden utilizar para realizar el seguimiento de los avances a lo largo del tiempo. Además, engloban un conjunto de cuestionarios de antecedentes para estudiantes, docentes, directores o padres a fin de reunir información sobre las condiciones socioeconómicas y demográficas, así como sobre las prácticas o experiencias que influyen en los resultados del aprendizaje (véase la **Tabla 4**). Por otra parte, algunas evaluaciones generan información de gran relevancia acerca del sistema educativo, en particular información que no suele estar disponible en las bases de datos internacionales sobre educación, por ejemplo, sobre los tipos de educación en la primera infancia; las políticas nacionales relativas al ingreso, el cambio de curso y los exámenes en la escuela; y el tiempo lectivo dedicado a una determinada materia.

Esta riqueza informativa permite a los países valorar la salud de su educación, detectar correlaciones entre los resultados (p. ej., las cualificaciones de los docentes, los recursos educativos o el acceso) y entender el modo en que interactúan las diversas variables para proporcionar una visión más clara de la dinámica de los sistemas educativos. En conjunto, los datos ofrecen información sobre esferas relacionadas con los logros académicos de los alumnos y su vinculación con la salud, la igualdad, la calidad de los insumos escolares, el grado de apoyo que reciben los alumnos en el hogar o el tipo y la ubicación de las escuelas. En este sentido, las LSLA desempeñan una función esencial en la medición de la equidad, uno de los principios clave que orientan la agenda Educación 2030.

21 Las actividades de distintas organizaciones encargadas de producir datos de evaluación reflejan ese potencial. Por ejemplo, la UNESCO y la IEA elaboraron un folleto para publicar los resultados del PIRLS 2016 y poner de relieve también de qué modo se pueden emplear las LSLA a la hora de llevar a cabo un seguimiento del progreso del ODS 4 mediante la reunión de datos sobre los logros en materia de aprendizaje y otras informaciones.

Si bien las evaluaciones de aprendizaje transnacionales tienden a incluir una gran variedad de cuestionarios contextuales sobre factores relativos a los hogares, las comunidades, las escuelas y los alumnos, esto no suele suceder en las evaluaciones nacionales del aprendizaje. La mayoría de las evaluaciones del aprendizaje se acompañan de un cuestionario para los alumnos (8 de cada 10) y la mitad de ellas, de un cuestionario para las escuelas o los docentes, mientras que solo el 20 % de ellas reúne información sobre las prácticas en los hogares, el entorno o el apoyo parental, todos ellos determinantes de gran importancia de los logros en materia de aprendizaje (IEU, 2019a). También existen otros factores que no se reflejan en ninguna evaluación. La información relativa a la salud y las discapacidades, el origen étnico y la condición de minoría a menudo se encuentra estrictamente regulada y no resulta sencillo obtenerla por medio de evaluaciones nacionales o transnacionales del aprendizaje.

Tabla 4. Información contextual reunida junto con los resultados del aprendizaje

Cuestionario para los estudiantes	Cuestionario para los docentes	Cuestionario para los padres	Cuestionario para las escuelas (completado por el director)
<ul style="list-style-type: none"> • Género, edad e idioma hablado en el hogar • Pertenencias en el hogar (p. ej., mesa de estudio, Internet, dormitorio, cantidad de libros y tipos de libros) • Prácticas en el hogar y apoyo parental (p. ej., uso de computadora, lectura por placer o actividades con los padres) • Salud y bienestar (faltas de asistencia a la escuela, sentimientos respecto a esta o nutrición adecuada en el hogar) • Su escuela (acoso, disfrute de las diferentes lecciones, impresiones sobre los docentes o uso de la biblioteca) • Medidas de las competencias de lectura y matemáticas comunicadas por los propios encuestados 	<ul style="list-style-type: none"> • Género, edad, nivel educativo, principales esferas de estudio y área de especialización • Experiencia docente y desarrollo profesional formal • Actitudes e impresiones con respecto a la escuela, sus recursos, el entorno y los factores asociados al éxito académico • Reflexiones sobre los comportamientos, las prácticas o los recursos que dificultan el aprendizaje en el aula • Prácticas y evaluación de la enseñanza • Participación de la familia en la escuela 	<ul style="list-style-type: none"> • Tipo y frecuencia de las actividades con los hijos (leer, cantar, contar, jugar, etc.) • Situación migratoria e idioma o idiomas que se hablan en el hogar • Participación de sus hijos en programas educativos para la primera infancia y duración de dicha participación • Edad de ingreso de sus hijos en la escuela primaria • Impresiones acerca del nivel de lectura, matemáticas o ciencias de sus hijos • Prácticas de estudio fuera de la escuela • Impresiones sobre la escuela de sus hijos • Prácticas de alfabetización de los padres, nivel educativo máximo y tipo de empleo 	<ul style="list-style-type: none"> • Experiencia y educación del director • Cantidad de alumnos inscritos y distribución en función de sus condiciones socioeconómicas • Ubicación geográfica de la escuela • Estado de la infraestructura académica • Comidas ofrecidas • Tiempo lectivo • Recursos y tecnología (p. ej., biblioteca, salas de informática, suministros e instalaciones para los alumnos con discapacidades) • Disciplina y seguridad de la escuela • Contribución de los padres y la comunidad a la escuela

Fuente: Autores. Para consultar más ejemplos de cuestionarios contextuales, véase Cresswell *et al.*, 2015; OCDE, 2017; IEU, 2016b; IEA, 2015; IEA, 2016; IEU, 2016b.

Nota: Ningún cuestionario contiene todos los elementos que se describen en esta tabla. El objetivo consiste en ejemplificar la gran variedad de información que se puede reunir y analizar junto con los niveles de aprendizaje a partir de preguntas contextuales adjuntas.

Con el fin de realizar un seguimiento de los compromisos contraídos en el plano normativo que se ponen en práctica a lo largo del tiempo, los cuestionarios adjuntos y otros datos complementarios sobre el sistema educativo ayudan también a recopilar información relevante en materia de políticas. Algunos de los datos que se pueden recopilar son los efectos de las políticas sobre los distintos idiomas de enseñanza, el modo en que evolucionan las normativas relativas a los docentes y el énfasis relativo de las sub-ámbitos evaluados en el plan de estudios. En resumen, los cuestionarios de antecedentes no solo pueden proporcionar información de gran valor sobre los resultados del aprendizaje de determinados grupos de población, sino también sobre los insumos relacionados con factores como los docentes, las aulas, el apoyo parental y los recursos escolares, que contribuyen a dicho aprendizaje. Esta riqueza informativa, a su vez, desempeña un papel crucial a la hora de guiar y orientar las iniciativas normativas a múltiples niveles (mundial, nacional y local) y, en última instancia, posibilitar la consecución del ODS 4.

Contribuciones a la nueva agenda educativa global

En los últimos decenios, han surgido diferentes iniciativas destinadas a prestar apoyo a la medición de los resultados del aprendizaje. En este apartado, se presenta una sinopsis de los diversos proyectos y programas que se han puesto en práctica recientemente a distintos niveles para contribuir al progreso y la consolidación de las LSLA.

Apoyo a las LSLA: iniciativas globales y regionales

Como se indicó anteriormente, las LSLA se están convirtiendo rápidamente en herramientas indispensables para el seguimiento de los resultados ligados al aprendizaje en la agenda Educación 2030. Este cambio ha contado con el respaldo de gobiernos donantes, organizaciones

Diversas organizaciones internacionales y regionales están promoviendo las LSLA para ayudar a mejorar los sistemas. Se han hecho esfuerzos significativos para ayudar a los países a generar evidencia y utilizarla en esferas clave para la mejora de los procesos y los resultados del aprendizaje.

internacionales y regionales, asociados para el desarrollo y otras entidades del sector privado que reconocen el enorme valor de estos instrumentos para la mejora de la educación. Estas organizaciones también entienden la importancia de los datos de las evaluaciones a la hora de elaborar estrategias nacionales para el desarrollo de la educación, así como la necesidad de aumentar las iniciativas en el plano nacional, regional y mundial para superar las grandes lagunas de datos en los niveles de aprendizaje registrados en todo el mundo (IEU, 2018a).

Diversas organizaciones internacionales y regionales están promoviendo las LSLA para ayudar a mejorar los sistemas. Entre las iniciativas consolidadas se encuentra,

por ejemplo, el programa de evaluación regional dirigido por el Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE) en la Oficina Regional de Educación de la UNESCO para América Latina y el Caribe. Desde 1997, se han llevado a cabo hasta tres evaluaciones regionales en esta zona, en los ámbitos de las matemáticas, la ciencia y la lectura. Más recientemente, se han hecho esfuerzos significativos para ayudar a los países a generar evidencia y utilizarla en esferas clave para la mejora de los procesos y los resultados del

aprendizaje (véase el **Cuadro 2**). Estas iniciativas y enfoques nuevos suponen un gran avance en nuestra concepción acerca del uso de los datos y la evidencia sobre el aprendizaje para la mejora de los sistemas educativos.

Cuadro 2. Iniciativas recientes en el plano regional y global para promover las evaluaciones de aprendizaje a gran escala

APOYO TÉCNICO

- El Instituto Internacional de Planeamiento de la Educación de la UNESCO (IIPE) está ampliando su apoyo a los Estados Miembros (principalmente en forma de investigación y capacitación) para fomentar el uso de los datos de las evaluaciones de aprendizaje en la planificación y la gestión de los sistemas educativos (IIPE, 2017).
- El Centro de Educación Universal de Brookings Institution presenta la iniciativa Optimización de las Evaluaciones para Todos (OAA) con el objeto de prestar apoyo a los países para mejorar las evaluaciones, la docencia y el aprendizaje de las competencias del siglo XXI, en colaboración con la iniciativa de la Red para el Monitoreo de la Calidad Educativa en la Región Asia-Pacífico (NEQMAP), en la Oficina de la UNESCO en Bangkok, y la iniciativa TALENT, en la Oficina de la UNESCO en Dakar (más información sobre estos proyectos en la parte inferior).²²
- La Asociación pro Desarrollo de la Educación en África (ADEA) creó la Red para la Evaluación del Aprendizaje en África (NALA) en 2016, un programa de investigación que trata de promover las evaluaciones relacionadas con el ODS 4, en especial las evaluaciones de aptitudes no cognitivas.
- El Análisis de los Sistemas Nacionales de Evaluación del Aprendizaje (ANLAS) desarrollado por ACER con el apoyo de la Alianza Mundial para la Educación tiene por objeto ayudar a los países a llevar a cabo un análisis exhaustivo de sus sistemas de evaluación.

REDES E INTERCAMBIO DE CONOCIMIENTOS

- La Red para el Monitoreo de la Calidad Educativa en la región Asia-Pacífico de la UNESCO (NEQMAP) trata de mejorar la capacidad de los Estados Miembros de emplear las evaluaciones del aprendizaje de los alumnos para fortalecer los sistemas educativos.
- En África Subsahariana, la Red de Docentes para la Transformación de la Enseñanza y el Aprendizaje (TALENT) trabaja también directamente con los Estados Miembros a fin de fortalecer las esferas críticas para mejorar el aprendizaje: los planes de estudios, las políticas y la capacitación de los docentes, y la evaluación de los resultados del aprendizaje.
- En 2016, el IIPE presentó el Portal de Aprendizaje con el objeto de ofrecer información actualizada, relevante y neutral sobre las cuestiones relativas al aprendizaje de la enseñanza primaria y secundaria.

FINANCIACIÓN

- La iniciativa Evaluación para el Aprendizaje de la Alianza Mundial para la Educación trabaja a nivel nacional con el fin de velar por que las inversiones en sistemas de evaluación del aprendizaje se fundamenten en un análisis de las necesidades, estén relacionadas con los procesos de elaboración de políticas públicas y se integren en los planes del sector de la educación. Hace poco, la Alianza Mundial para la Educación lanzó el mecanismo Intercambio de Conocimientos e Innovaciones (KIX), que también se centra específicamente en respaldar los sistemas de evaluación del aprendizaje. El mecanismo KIX se basa en la iniciativa Evaluación para el Aprendizaje (A4L), que se puso en práctica en 2017 y se centra en el fomento de la capacidad y la prestación de apoyo a los países a fin de fortalecer sus sistemas de evaluación del aprendizaje.

Fuente: Autores.

Nota: La selección de iniciativas presentada aquí no pretende ser exhaustiva, sino ejemplificar el énfasis de las iniciativas regionales y mundiales en las LSLA.

22 Si desea obtener más información sobre la iniciativa OAA, consulte la página https://www.brookings.edu/wp-content/uploads/2017/12/overview_optimizing-assessment-for-all.pdf.

A lo largo de los últimos años, las organizaciones internacionales han empezado a centrarse cada vez más en los datos, contribuyendo así a la creación de sólidos sistemas de reunión de datos. Incluso antes de la aprobación de la Agenda 2030, se multiplicaron las iniciativas en apoyo de la medición de los resultados del aprendizaje con diversas opciones encaminadas a establecer métricas mundiales para la lectura y la aritmética. Entre ellas, se incluye la Comisión Especial sobre Métricas de los Aprendizajes (2012-2016)²³, la Alianza para las Métricas del Aprendizaje (2014)²⁴ y el Grupo de Estudio sobre la Medición de los Resultados del Aprendizaje del Centro para el Desarrollo Mundial (2013)²⁵. En 2016, se puso en marcha la Alianza Mundial para el Seguimiento del Aprendizaje (GAML), dirigida por el Instituto de Estadística de la UNESCO (IEU), con el objeto de consolidar diferentes procesos. La GAML ofrece un espacio para dialogar sobre perspectivas políticas y conocimientos técnicos a la hora de establecer estrategias, metodologías y herramientas destinadas a llevar a cabo un seguimiento de los resultados del aprendizaje en el plano mundial (véase el **Cuadro 3**). La creación de la GAML constituye un importante paso hacia una mayor sinergia al promover la colaboración entre las diversas partes interesadas, muchas de las cuales compiten por una cuota del “mercado” de la medición de los resultados del aprendizaje.

Cuadro 3. La Alianza Mundial para el Seguimiento del Aprendizaje

La Alianza Mundial para el Seguimiento del Aprendizaje (GAML) es una iniciativa de múltiples interesados dirigida por el IEU con el objetivo de prestar apoyo a las estrategias nacionales para medir el aprendizaje y posibilitar la elaboración de informes internacionales sobre el marco de indicadores del ODS 4. La GAML reúne a Estados Miembros de las Naciones Unidas, expertos técnicos internacionales y asociados para la ejecución (Estados Miembros, donantes, sociedad civil, organismos de las Naciones Unidas y sector privado) para mejorar la evaluación del aprendizaje en todo el mundo. La labor de la GAML se organiza alrededor de equipos de trabajo que abordan las cuestiones ligadas a los datos sobre resultados del aprendizaje comparables a nivel mundial.

La GAML está desarrollando un consenso global respecto a la presentación de información y la medición del aprendizaje a nivel internacional, sobre la base de las prácticas y los enfoques metodológicos que mejor se adaptan al cometido. Está desarrollando así: 1) todos los indicadores del ODS 4 relacionados con el aprendizaje y las aptitudes, así como las herramientas y normas metodológicas necesarias para garantizar la comparabilidad en el plano mundial; 2) normas, directrices y herramientas que ayuden a los países a fortalecer la implementación de sus evaluaciones y a evaluar su calidad; y 3) herramientas de fomento de la capacidad que ayuden a los países a reunir, analizar y utilizar los datos de evaluación del aprendizaje. El trabajo metodológico de la GAML sirve de fundamento para otros procesos globales.

Fuente: Adaptación del sitio web de la GAML del IEU: <http://gaml.uis.unesco.org/>.

23 Para obtener más información sobre la Comisión Especial sobre Métricas de los Aprendizajes: <https://www.brookings.edu/product/learning-metrics-task-force/>.

24 Para obtener más información sobre la Alianza para las Métricas del Aprendizaje: http://uis.unesco.org/sites/default/files/documents/learning-metrics-partnership-a-capacity-support-and-policy-strengthening-initiative-to-develop-and-use-common-learning-metrics-mathematics-reading-2014-en_1.pdf.

25 Para obtener más información sobre el Grupo de Estudio y sus productos: <https://www.cgdev.org/publication/ft/schooling-not-education-using-assessment-change-politics-non-learning>.

Una tendencia mundial al alza que es también motivo de preocupación

Las LSLA son instrumentos ampliamente aceptados que ayudan a los países a garantizar la *equidad* y la *calidad* de la educación. Por consiguiente, se espera que el aumento y la difusión de las LSLA que se ha observado en todo el mundo (en el plano nacional y transnacional) continúe, especialmente en vista de que, a lo largo del último decenio, los agentes globales y regionales de la esfera educativa también han replanteado sus identidades y agendas de trabajo en el ámbito educativo para incorporar las LSLA (Addey, 2018), al mismo tiempo que son cada vez más los asociados para el desarrollo que animan a los países a participar en ese tipo de iniciativas.

A consecuencia de estas tendencias, hoy en día el debate mundial relativo a las políticas educativas fundamentalmente gira en torno a la medición y las comparaciones de los resultados educativos. No obstante, diferentes analistas han mostrado su preocupación por el hecho de que la mayor atención prestada a la medición de los resultados del aprendizaje pueda influir en la política y la práctica educativa de un modo que podría resultar perjudicial para el programa general del ODS 4. Algunas de estas preocupaciones se derivan de las características inherentes al diseño de las LSLA, mientras que otras de ellas se deben a los posibles análisis, interpretaciones y usos correctos (o incorrectos) de los datos reunidos. En los capítulos siguientes, se analizan todas estas limitaciones y se hace especial hincapié en las repercusiones para el avance hacia el compromiso mundial de un aprendizaje equitativo, efectivo y relevante para todos.

Diferentes analistas han mostrado su preocupación por el hecho de que la mayor atención prestada a la medición de los resultados del aprendizaje pueda influir en la política y la práctica educativa de un modo que podría resultar perjudicial para el programa general del ODS 4. Algunas de estas preocupaciones se derivan de las características inherentes al diseño de las LSLA, mientras que otras de ellas se deben a los posibles análisis, interpretaciones y usos correctos (o incorrectos) de los datos reunidos.

3. Valorar más de lo que puede medirse: las limitaciones del diseño de las evaluaciones de aprendizaje a gran escala

En el presente capítulo, se analizan las consecuencias imprevistas de las evaluaciones de aprendizaje a gran escala (LSLA) que se derivan directamente de su diseño. El apartado trata sobre las cuatro dimensiones principales que corren el riesgo de verse perjudicadas por la propia naturaleza y el propio diseño de las LSLA: la conceptualización de la educación y el aprendizaje; el entendimiento general de la calidad de la educación; la variedad de conocimientos, aptitudes, actitudes y valores que se toman en consideración; y la atención prestada a la inclusividad y el respeto de la diversidad.

Limitando la conceptualización de la educación y el aprendizaje

En este primer apartado, se estudia cómo un énfasis excesivo en las LSLA puede constreñir la conceptualización de la educación y el aprendizaje. El apartado comienza con una sinopsis de los diversos propósitos de la educación y las múltiples dimensiones del aprendizaje. Así, analiza cómo, a consecuencia de un enfoque centrado en las mediciones cuantitativas y la comparabilidad a nivel del sistema, a menudo se pasan por alto o se descuidan otras modalidades de evaluación más adecuadas para reflejar ciertos ámbitos del aprendizaje y, lo que es más importante, se desvía la atención de la visión global de la educación y sus propósitos generales.

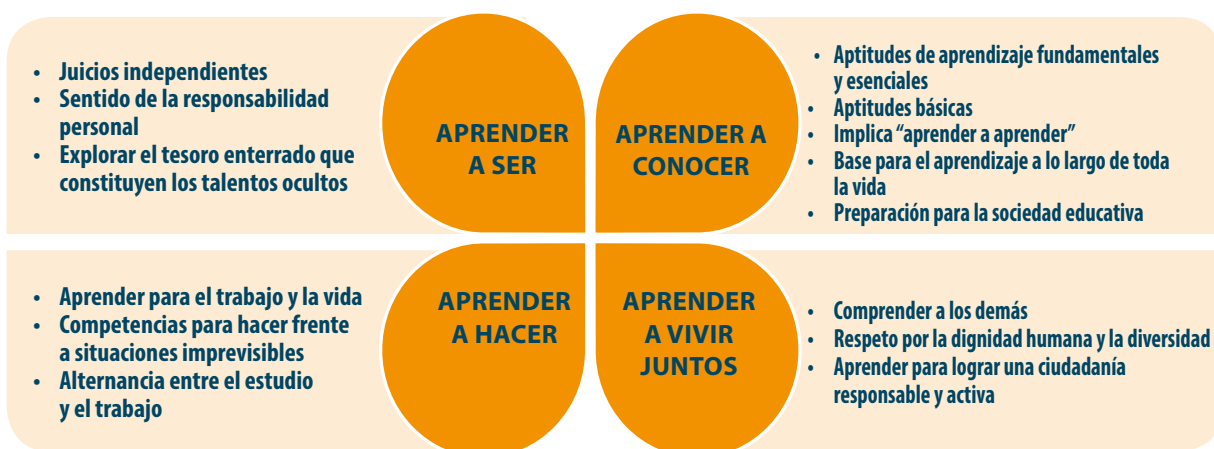
Recordando los diversos propósitos de la educación

Cabe recordar que la educación cumple múltiples propósitos tanto para las personas como para las comunidades. La educación no conlleva solo la adquisición de aptitudes básicas y profesionales, conocimientos y competencias generales, sino también el desarrollo de valores de respeto a la vida, a la dignidad humana y a la diversidad cultural, necesarios para la armonía social en un mundo caracterizado por la diversidad (UNESCO, 2015a). Se trata de tomar conciencia de las decisiones que se adoptan a diario para llevar una vida saludable y plena, contribuyendo al mismo tiempo a los componentes social, económico y ambiental del desarrollo sostenible. La educación reviste una gran relevancia para las distintas dimensiones del desarrollo humano, tanto a nivel personal como colectivo, y se encuentra estrechamente interrelacionada con dichas dimensiones. Por consiguiente, implica el equilibrio y la integración de los múltiples propósitos sociales, económicos, culturales, éticos, ambientales, espirituales y políticos de la educación.

Reafirmando la relevancia de los cuatro pilares

El aprendizaje consiste en un proceso complejo de adquisición o desarrollo de conocimientos, aptitudes y actitudes con una base de valores subyacentes (UNESCO, 2017c). Los cuatro pilares del aprendizaje (UNESCO, 1996) recogen un enfoque integrado para el aprendizaje: aprender a *conocer*, a *hacer*, a *ser* y a *vivir juntos* (véase la **Figura 1**). Aprender a *conocer* consiste en aprender aptitudes básicas y transferibles que sirven de fundamento para el aprendizaje a lo largo de toda la vida, así como de preparación para la participación activa en la sociedad educativa. Aprender a *hacer* implica aprender a aplicar los conocimientos y las aptitudes adquiridos a situaciones conocidas o imprevisibles en el trabajo y en la vida. Aprender a *ser* supone aprender a desarrollar la propia personalidad, además de la capacidad de formular juicios independientes y de tener sentido de la responsabilidad. Aprender a *vivir conjuntos* conlleva la capacidad de entender a los demás y de tener respeto por la dignidad humana, la diversidad y las culturas a fin de lograr una ciudadanía responsable y activa.

Figura 1. Los cuatro pilares de la educación



Fuente: Adaptación de UNESCO, 1996.

Evaluaciones más allá de las mediciones, aprendizaje más allá de las evaluaciones

Los contenidos de los planes de estudios y los resultados previstos de las prácticas de aprendizaje son muy amplios. Sin embargo, las LSLA, por su diseño, suelen limitarse a un número reducido de esferas de aprendizaje que se pueden medir con facilidad, lo que permite utilizarlas para hacer comparaciones al nivel del sistema o entre distintos países. Estas áreas mensurables representan "aptitudes básicas" o conocimientos generales fundamentales, como la lectura y las matemáticas (y, en ocasiones, las ciencias), que se consideran

Los contenidos de los planes de estudios y los resultados previstos de las prácticas de aprendizaje son muy amplios. Sin embargo, las LSLA, por su diseño, suelen limitarse a un número reducido de esferas de aprendizaje que se pueden medir con facilidad, lo que permite utilizarlas para hacer comparaciones al nivel del sistema o entre distintos países, y/o a las denominadas "aptitudes básicas" o conocimientos generales fundamentales.

primordiales para fomentar las aptitudes y competencias, así como para seguir aprendiendo en estos ámbitos y en otros. Por tanto, las LSLA tienden a dejar de lado una gran variedad de dimensiones del aprendizaje que son importantes para los múltiples propósitos de la educación²⁶. Se puede decir que esta cuestión no es intrínsecamente problemática. No obstante, en los contextos donde exista una dependencia excesiva de las LSLA, corremos el riesgo de valorar lo que medimos en lugar de medir el aprendizaje que valoramos (Biesta, 2009).

Esto es así en particular en vista de que el debate mundial en materia de educación tiende cada vez más a equiparar la evaluaciones a las mediciones cuantitativas orientadas a la comparación de los sistemas, al tiempo que obvia otros tipos de evaluación. Las evaluaciones son esenciales para el aprendizaje. No obstante, es cuestionable si es posible o incluso deseable registrar el aprendizaje social, cívico y personal, de gran importancia, a través de métodos cuantitativos tradicionales. Por ejemplo, las competencias clave en materia de comunicación, colaboración, creatividad, ética y relaciones sociales dependen en gran medida de la cultura y están estrechamente vinculadas con la personalidad. Por lo tanto, para evaluar correctamente ese tipo de competencias es necesario que los alumnos demuestren sus actitudes y percepciones, y no solo sus conocimientos y su comprensión. Estas se pueden registrar mejor mediante distintos tipos de evaluaciones, como las evaluaciones de portafolio, la observación o incluso distintas clases de evaluaciones formativas cuyo objetivo no sea la comparabilidad. Aun así, prácticamente no hay cabida para este tipo de métodos de evaluación alternativos a fin de llevar a cabo un seguimiento del aprendizaje en los sistemas.

Las evaluaciones son esenciales para el aprendizaje. No obstante, es cuestionable si es posible o incluso deseable registrar el aprendizaje social, cívico y personal, de gran importancia, a través de métodos cuantitativos tradicionales

Por último, debemos ser conscientes de que, aunque el aprendizaje es ubicuo, los debates sobre su evaluación tienden a centrarse en la parte del aprendizaje con un fin determinado y organizado, además de en los resultados que se pueden medir. Pese a que el aprendizaje puede ser deliberado, intencionado, con un fin determinado y organizado, con frecuencia suele estar mucho menos institucionalizado y puede tener lugar tanto de modo intencionado como accidental en diversas actividades de la vida cotidiana, en el lugar de trabajo o en programas educativos no académicos, en la comunidad local o de forma autodidacta, a cargo de la familia o de la sociedad (UNESCO, 2015a). Al hacer un especial hincapié en las LSLA en el debate sobre la educación, existe el riesgo de reducir la noción de aprendizaje al aprendizaje mensurable.

Limitando nuestra concepción de una buena educación

A raíz de esta dinámica, es posible que los organismos internacionales y los gobiernos nacionales, a través del aumento de su participación en las LSLA y de la atención prestada a estas, estén contribuyendo a crear una realidad en la que estas evaluaciones determinan en gran

26 Esto es lo que sucedió con las evaluaciones centradas en los aspectos utilitarios del aprendizaje y, en especial, en su contribución al desarrollo económico, particularmente el PISA. Tradicionalmente, otros programas de evaluación se han caracterizado por un enfoque más comprehensivo y, por consiguiente, prestan atención a un mayor número de dimensiones del aprendizaje. Sin embargo, por motivos que están fuera del alcance de esta publicación, los programas de evaluación existentes presentan una convergencia cada vez mayor con relación a las esferas de medición, lo que se traduce en un mayor hincapié en la aritmética y la alfabetización.

parte nuestra concepción de lo que resulta valioso en el ámbito de la educación y los objetivos conexos (Hamilton, 2017). En última instancia, esta dinámica limita nuestra concepción general de lo que constituye una buena educación (Biesta, 2009) y probablemente aleja los debates de dimensiones educativas de importancia que no se deben pasar por alto. También es esencial apuntar que estos cambios no solo se producen a nivel discursivo. Como se explica en el capítulo siguiente, en realidad es probable que estas transformaciones tengan efectos concretos al institucionalizarse en forma de textos normativos con la facultad de transformar las identidades y las prácticas de los docentes, los planes de estudios formales y ocultos que se imparten en las escuelas, y las prioridades de las diversas partes interesadas en la educación.

Limitando nuestra concepción de la calidad de la educación

En este apartado, se analiza cómo un aumento del enfoque centrado en los resultados del aprendizaje como dimensión clave de la calidad de la educación puede desviar la atención lejos de otros componentes relevantes (p. ej., el impacto a nivel social, o los entornos de las escuelas y las aulas, y los procedimientos de enseñanza y aprendizaje), o de la influencia de factores macro que afectan a los resultados del alumnado. El apartado comienza con una presentación breve de la naturaleza multidimensional del concepto de calidad de la educación y prosigue con un debate sobre el enfoque equilibrado que se necesita para englobar los distintos componentes e indicadores de la calidad de la educación.

Calidad de la educación: un concepto multidimensional

Los enfoques tradicionales que miden la calidad de la educación se basaban en indicadores indirectos de la educación impartida y en insumos que afectan al progreso de la enseñanza y el aprendizaje, como la ratio de alumnos por maestro, el porcentaje de docentes capacitados o el gasto por alumno como porcentaje del producto interno bruto (PIB)²⁷. Recientemente, los resultados del aprendizaje han ido adquiriendo importancia como indicador esencial para medir la calidad de la educación. En el *Informe de Seguimiento de la Educación en el Mundo (GEM)*, por ejemplo, se propuso recientemente un marco revisado para entender la calidad de la educación (UNESCO, 2017d), que incluye un importante énfasis en los resultados del aprendizaje para las personas y las sociedades conforme al desempeño en las LSLA.

Sin embargo, tal como se subraya en el Informe GEM, los resultados individuales no pueden entenderse de manera aislada ni tampoco son la única medición válida de la calidad de la educación. También es necesario tomar en consideración otras dimensiones distintas de los resultados del aprendizaje. Algunas de ellas son el *impacto a nivel social* (p. ej., el fomento de comportamientos asociados al desarrollo sostenible desde el punto de vista social y ambiental, la ciudadanía mundial, la convivencia y la diversidad), así como el *entorno en las escuelas y las*

27 Estos enfoques se sintetizan en Tawil *et al.* (2012). Hacen referencia, por ejemplo, al modelo de Santiago de la UNESCO (Oficina de la UNESCO en Santiago de Chile, 2007), que propone cinco dimensiones de la calidad de la educación con las que pretende abarcar las perspectivas de los diversos interesados del sistema educativo; el marco analítico formulado por Tikly y Barret (2007), que demanda un análisis de los contextos histórico, socioeconómico y político para entender cómo funcionan los sistemas educativos; y el modelo de “urdimbre” de la calidad en la educación de Nikel y Lowe (2010), que propone siete dimensiones conceptuales tramadas para poner de relieve que la calidad de la educación es muy parecida a una “urdimbre”, es decir, es más resistente cuando está “estirada” o cuando se mantiene en tensión.

aulas. Esta última dimensión engloba el proceso de enseñanza y aprendizaje; la presencia de docentes adecuadamente capacitados y motivados, junto con un proceso riguroso de enseñanza; un sólido liderazgo de la escuela; la participación comunitaria y la gobernanza democrática; y las estructuras y los insumos materiales necesarios²⁸.

En líneas más generales, para entender los niveles de los resultados del aprendizaje o cómo cambian con el tiempo, es necesario tomar en consideración *factores generales* o *macro*, por ejemplo, los contextos económicos, políticos y sociales y la organización del sistema educativo (es decir, políticas públicas, financiación del sistema, plan de estudios, normas y rendición de cuentas, contratación e incentivos, y desarrollo profesional). De lo contrario, se obtendría inevitablemente un análisis y una perspectiva sesgados. Por tanto, la mejora de la calidad de la educación no conlleva necesariamente una mejora de los resultados del aprendizaje, ya que la variación del aprendizaje de los estudiantes no se atribuye únicamente a factores ligados a las escuelas, sino también a características de las familias y comunidades (Ungerleider, 2006). Aunque muchas LSLA tienen por objeto precisamente conocer mejor los efectos de esta clase de factores, a menudo la sociedad, los políticos y los profesionales de la educación no prestan atención al carácter fundamental de estas variables, en particular si los resultados difundidos se centran en perspectivas comparativas y, en el caso de las evaluaciones transnacionales, hacen hincapié en la clasificación nacional con respecto a otros países. A pesar de que las LSLA son necesarias para realizar un seguimiento del desempeño de los sistemas educativos y mejorar el aprendizaje y la equidad, su potencial aumenta cuando se tienen en cuenta junto con otros factores diversos relacionados con las escuelas y ajenos a estas.

Evaluaciones: un componente necesario pero insuficiente para reflejar la calidad de la educación

Comprender las múltiples dimensiones adicionales relacionadas con el alumno, el docente, la escuela, el sistema educativo y los contextos económico, social y político es primordial para evaluar la calidad de la educación (Colclough, 2012; UNESCO, 2016b; Zadjia, 2014). Los datos sobre estas dimensiones no siempre se generan mediante las LSLA, sino a través de otras fuentes, como las encuestas administrativas o de los hogares, ya que los datos reunidos por medio de cuestionarios de antecedentes no son suficientes para todos los casos. Por tanto, los datos de las LSLA que se toman en consideración y se analizan junto con diversos datos adicionales generados en los países reflejarán mejor las diferentes dimensiones de la educación de calidad.

Generalmente son los investigadores y los especialistas de diferentes organismos gubernamentales o consorcios de evaluación quienes suelen llevar a cabo este tipo de análisis. Sin embargo, a menudo otros informes menos especializados, como los que aparecen en los medios de comunicación, restan o directamente no dan ninguna importancia a estas relaciones. Además, parece que los enfoques ponderados y multidimensionales son cada vez más inusuales, a medida que sigue aumentando la relevancia de las LSLA. De hecho, sus resultados cuentan con

28 Unterhalter y Brighouse (2007) expresan una idea similar, ya que, basándose en la perspectiva del enfoque de las capacidades, señalan que los resultados del aprendizaje resultan inadecuados o insuficientes para reflejar las dimensiones clave de la calidad de la educación ajenas al propósito fundamental, incluidos los valores intrínsecos y posicionales de la educación.

una aceptación cada vez mayor como medición completa de la educación de calidad (Hanushek y Woessmann, 2015; Komatsu y Rappleye, 2017) y también han sido considerados como el componente principal (ACER-IEU, 2017) o como una base de referencia esencial (Friedman *et al.*, 2016) de la educación de calidad. Aunque, en líneas generales, esto pone de relieve una preocupación por el aprendizaje como resultado de gran importancia de la escolarización, existe el riesgo de que la calidad de la educación pasee a equipararse únicamente a la mejora de los resultados del aprendizaje conforme al seguimiento de las LSLA (Sayed *et al.*, 2018; Smith, 2016).

En suma, los niveles de los resultados del aprendizaje que se miden mediante LSLA constituyen una dimensión necesaria, pero no la única que hace falta para medir la calidad de la educación.

El uso de los resultados de las evaluaciones del aprendizaje como método único para evaluar la calidad de la educación conlleva múltiples repercusiones. Uno de sus riesgos es la desatención de esferas importantes de la educación de calidad, como luchar por la equidad; garantizar entornos de aprendizaje adecuados y eficaces; contar con docentes cualificados y ampliar sus capacidades; velar por la relevancia de los planes de estudios; y ofrecer un liderazgo sólido en las escuelas. Por otra parte, puede llevar a descuidar otras dimensiones sociales, económicas y contextuales de la calidad de la educación. En suma, los niveles de los resultados del aprendizaje que se miden mediante LSLA constituyen una dimensión necesaria, pero no la *única que hace falta para medir* la calidad de la educación. Se puede incluso cuestionar si la calidad se puede valorar de un modo adecuado sin hacer referencia a estas otras dimensiones sociales, económicas y contextuales.

Restringiendo los aspectos que valoramos

Como se explicó anteriormente, las LSLA pueden provocar la limitación de los múltiples propósitos de la educación y las dimensiones del aprendizaje. En este apartado, se profundiza en uno de los procesos que causa esta limitación, a saber, la cantidad reducida de ámbitos en los que se centran la mayor parte de las LSLA. En este apartado, se analizan en primer lugar las distintas modalidades a través de las que se produce esta reducción curricular y, a continuación, se incluyen ejemplos de iniciativas encaminadas a incorporar una mayor variedad de conocimientos, aptitudes, actitudes y valores en las LSLA, junto con los obstáculos conexos que se encuentran. Finalmente, la sección reflexiona sobre la conveniencia de este tipo de iniciativas y los posibles riesgos que entrañan.

Ámbitos y sub-ámbitos limitados

Como ya se ha indicado, se puede considerar que el número limitado de ámbitos y capacidades en los que se centran las LSLA es un aspecto intrínseco a su diseño. Estas evaluaciones se concentran necesariamente en un número limitado de materias y capacidades, en concreto, en esferas del aprendizaje que se pueden medir con facilidad y permiten realizar comparaciones al nivel de sistema o entre distintos países, o que se consideran aptitudes básicas para continuar aprendiendo en otras áreas.

Por consiguiente, algunas personas defienden que el enfoque de las LSLA es demasiado limitado en dos sentidos. En primer lugar, no evalúan adecuadamente una gran variedad de conocimientos en un amplio espectro de ámbitos, como las competencias en la esfera de las artes, los medios de comunicación y la información en las cibernaciones actuales, y tampoco evalúan competencias generales que se deben desarrollar en las escuelas, como la ciudadanía o la responsabilidad ambiental. En segundo lugar, dentro de los ámbitos que sí se evalúan, las LSLA no prestan suficiente atención a todos los conocimientos disponibles. Por ejemplo, a la hora de evaluar las competencias lingüísticas, a menudo las LSLA no evalúan sub-ámbitos como la valoración de la literatura, la expresión escrita y la comprensión oral, pese a que son objetivos importantes de la disciplina (Ungerleider, 2006).

¿Más allá de la aritmética y la alfabetización?

Así, el diseño de las LSLA tiende a pasar por alto los distintos propósitos de la educación y las dimensiones del aprendizaje y, de manera implícita, los principios esenciales de los compromisos de la agenda Educación 2030. Esto se debe tanto al contenido de las evaluaciones como a su proceso de administración y al formato de pregunta que se utiliza a la hora de diseñarlas. Los elementos de las pruebas, de opción múltiple o respuesta corta, se conciben de modo que se puedan puntuar fácilmente en función de un conjunto definido de criterios, lo que limita las posibilidades de evaluar percepciones, actitudes y juicios. De manera similar, la mayoría de las evaluaciones no tratan necesariamente de reunir datos relevantes para los estilos de vida saludables o los valores y las actitudes que respaldan la democracia, la protección de los derechos humanos, el respeto de la diversidad cultural, la ciudadanía responsable y activa, y la sostenibilidad ambiental²⁹.

En este apartado, reflexionamos sobre algunos de los conocimientos, ámbitos, aptitudes, actitudes y valores que, pese a ser elementos clave de la agenda Educación 2030, solo se reflejan de manera parcial (y con grandes dificultades) en las LSLA. Una panorámica de algunos elementos de prueba utilizados en estos instrumentos de evaluación ayudará a ilustrar la dificultad de incorporar una gran variedad de competencias (es decir, ámbitos de aprendizaje, conocimientos, aptitudes, actitudes y valores) a las prácticas de evaluación habituales.

Estilos de vida saludables: ámbitos con un análisis insuficiente y retrocediendo

La educación se considera una de las formas más poderosas de mejorar la salud de las personas y velar por que los beneficios se transmitan a las generaciones futuras. A su vez, la salud es un facilitador y un determinante clave del aprendizaje (UNESCO, 2016b; 2016c). Sin embargo, la mayoría de las LSLA no apoyan de manera explícita una educación encaminada a lograr estilos de vida saludables. En lo que a las evaluaciones nacionales se refiere, la salud, la nutrición y el bienestar de los niños, así como de sus padres o tutores, se mide con mayor frecuencia en el desarrollo de la primera infancia, ya que se considera una dimensión clave para garantizar que los niños están aprendiendo y están en buen camino para la enseñanza primaria (UNESCO *et al.*, 2017). Apenas hay indicios de que la noción de estilos de vida saludables sea un componente esencial de las LSLA centradas en niños de primaria y secundaria, o incluso en adultos. Esto

²⁹ Las dificultades ligadas a la evaluación de estos ámbitos no se pueden separar del hecho de que casi nunca constituyen materias independientes en los planes de estudios nacionales, sino que se suelen integrar en los programas de estudios como temas transversales.

se aplica especialmente a las evaluaciones desarrolladas en el plano nacional. Si bien cabe reconocer que los cuestionarios de antecedentes suelen reunir datos sobre las prácticas y los hábitos de los estudiantes con respecto a la salud o la nutrición, por lo general solo contienen preguntas destinadas a determinar si los alumnos están o no *en condiciones* de aprender, es decir, si tienen acceso a los servicios de salud adecuados, si gozan de una nutrición apropiada y suficiente en la escuela, etc. En general, no se reúne ninguna información relativa a sus enfoques respecto a la prevención de enfermedades o sus opiniones sobre comportamientos como los hábitos de salud y alimentarios o la exposición al riesgo sexual. Asimismo, aunque es posible que algunos elementos estén concebidos para abordar cuestiones de salud, el objetivo de las tareas suele ser medir aptitudes concretas, como la comprensión lectora o la resolución de problemas en múltiples pasos.

Por ejemplo, el ítem de del PISA de 2015 Ciencias “Correr en días de calor” (OECD, 2015)³⁰, incluye cinco preguntas sobre la interacción entre la salud y el medio ambiente. Este elemento trata de evaluar diferentes capacidades: interpretar la presentación; manipular una variable y comparar los resultados de dos pruebas; deducir la función del sudor a partir del texto escrito y la simulación; recurrir a conocimientos procedimentales para explicar cómo justifican su respuesta los datos que han reunido; y desarrollar una hipótesis realizando una extrapolación más allá de los datos que se pueden recopilar directamente mediante la simulación. No se espera que el alumno tenga ningún conocimiento anterior sobre el tema para responder a las preguntas, ya que las respuestas se pueden deducir de los estímulos. De la misma forma, no se pretende evaluar las decisiones, las medidas ni los juicios relativos a la actividad física en días de calor. Por consiguiente, los cuestionarios de prueba del proyecto PISA utilizan materiales relacionados con la salud para evaluar las competencias científicas y de resolución de problemas, pero el estímulo ofrece todas las pistas, de manera que se excluye el elemento del juicio de la evaluación.

No obstante, se han llevado a cabo algunas iniciativas encaminadas a hacer un mayor hincapié en la noción de los estilos de vida saludables. Por ejemplo, en respuesta a las elevadas tasas del virus de la inmunodeficiencia humana/síndrome de inmunodeficiencia adquirida (VIH/sida) en África Subsahariana, el Consorcio de África Meridional y Oriental para la Supervisión de la Calidad de la Educación (SACMEQ) de 2007 incluyó un módulo sobre el VIH y el sida para determinar las actitudes de los niños respecto a las personas que viven con la enfermedad, sus percepciones acerca de la evaluación del VIH y el sida, y su fuente preferencial de información sobre la enfermedad³¹. Se estudió también el grado en que la formación del profesorado aborda cuestiones relativas al VIH/sida, centrales para cualquier estrategia efectiva de promoción de la prevención (Dolata, 2011).

Educación para la ciudadanía mundial y para el desarrollo sostenible: ámbitos cada vez más definidos y evaluados

La educación para el desarrollo sostenible (EDS) y para la ciudadanía mundial constituye una de las dimensiones clave de la agenda Educación 2030. Ya en 1974, la Conferencia General de la

30 El ítem de PISA de 2015 Ciencias “Correr en días de calor” puede encontrarse en el siguiente link: <http://www.oecd.org/pisa/PISA2015Questions/platform/index.html?user=&domain=SCI&unit=S623-RunningInHotWeather&lang=esp-esp>

31 Sin embargo, cabe destacar que el módulo relativo al VIH y el sida no se incluyó en el último ciclo del SACMEQ (2013-2014).

UNESCO aprobó la *Recomendación sobre la educación para la comprensión, la cooperación y la paz Internacionales y la educación relativa a los derechos humanos y las libertades fundamentales*. La Recomendación de 1974 alentó a los Estados a tomar las medidas institucionales y pedagógicas necesarias para incorporar determinados principios (es decir, el respeto por la diversidad cultural, la promoción de la paz y la resolución no violenta de conflictos, la justicia, el cuidado del planeta, etc.) en sus programas educativos (UNESCO, 1974).

Si bien este tipo de iniciativas pueden revestir una gran importancia, el diseño de medidas encaminadas a evaluar los conocimientos, las aptitudes, los valores y las actitudes conexos es una tarea compleja. Esto se debe en particular a que las herramientas existentes se limitan a menudo a la información proporcionada por los mismos interesados mediante auto-informes, y, por lo general, se ven afectados por un sesgo de deseabilidad social, ya que los participantes suelen tratar de responder de un modo que les parezca socialmente aceptable. En las evaluaciones transnacionales, puede resultar difícil reflejar las normas relevantes desde el punto de vista cultural que son específicas de determinados países o regiones y conservar al mismo tiempo la comparabilidad de los datos reunidos entre las distintas naciones, lo que agrava todavía más el problema. Hoy en día, parece que los conocimientos técnicos para poner remedio a este problema se encuentran en una etapa de desarrollo. Además, no está claro si la EDS se incluirá en las evaluaciones nacionales del aprendizaje en un futuro próximo. En Europa, pese a que la dimensión del conocimiento de la ciudadanía mundial suele formar parte de las evaluaciones nacionales, existen menos probabilidades de que se aborden las actitudes y aptitudes de los estudiantes (De Coster y Sigalas, 2017).

Reflejar la educación para la ciudadanía en las evaluaciones nacionales resulta especialmente complicado, dado que engloba una gran variedad de enfoques de enseñanza y aprendizaje, desde la impartición de asignaturas separadas o contenidos transversales hasta las actividades extracurriculares y la participación en la vida escolar (De Coster y Sigalas, 2017). No obstante, se están poniendo en práctica distintas iniciativas para lograrlo. Diferentes evaluaciones transnacionales del aprendizaje como la evaluación mundial de competencias PISA, de la OCDE, y el Estudio Internacional sobre Educación Cívica y Ciudadana (ICCS), de la IEA, están tratando de evaluar los componentes de la EDS y la educación para la ciudadanía. Asimismo, el ICCS, que se ha puesto en práctica en 38 países, está diseñado como una evaluación no curricular destinada a reflejar cómo se preparan la juventud para asumir sus funciones como ciudadanía (IEA, 2016b). El estudio ofrece información sobre los conocimientos de los estudiantes y su entendimiento de conceptos y cuestiones relacionados con la educación cívica y ciudadana, así como sus creencias, actitudes y comportamientos. El **Cuadro 4** presenta un ejemplo del enfoque adoptado por el ICCS para evaluar los conocimientos sobre la diversidad religiosa, además de las actitudes y la inclinación hacia la tolerancia. La IEA también promovió la elaboración de informes regionales que analizaran los contextos de la educación cívica y ciudadana, así como las percepciones de los estudiantes de instituciones públicas y del Gobierno, sus opiniones sobre la coexistencia pacífica y sus puntos de vista acerca de la cohesión social y la diversidad³².

32 Para reflejar estas dimensiones, los alumnos disponen de espacio para escribir y expresar sus opiniones, lo que supone un nuevo desafío a la hora de establecer la puntuación de las evaluaciones.

Cuadro 4. Coexistencia respetuosa: elemento de prueba del Estudio Internacional sobre Educación Cívica y Ciudadana de 2009

A continuación aparece una calcomanía que se puede comprar por Internet.



La calcomanía está formada por símbolos que representan diferentes maneras de ver el mundo. Los símbolos se han agrupado de manera que se parezcan a la palabra inglesa “coexist”, en español “coexistir”, que significa “vivir juntos”.

CI2COM1

- 1 ¿Cuál es el propósito **más probable** de esta calcomanía?
- ☐ mostrar que las distintas formas de pensar son todas iguales
 - ☐ mostrar que las personas deben reflexionar cuidadosamente sobre sus creencias
 - ☐ mostrar que la gente puede aceptar a otras personas incluso si sus creencias son distintas
 - ☐ mostrar que las personas que tienen distintas formas de ver el mundo no pueden vivir felices juntas

Fuente: Brese *et al.* 2011, pág. 7 (reproducido con autorización).

En el Sur Global, el programa SEA-PLM es actualmente la única evaluación transnacional que cuenta con un módulo destinado a medir las actitudes y los comportamientos ligados a la ciudadanía mundial. Dicho módulo se centra en actitudes y valores, antes que en resultados cognitivos, comportamientos y aptitudes, y se espera que refleje valores universales, además de valores específicos de las naciones de Asia Sudoriental. Por consiguiente, este marco de evaluación recientemente desarrollado reconoce los desafíos que conllevan las esferas de evaluación innovadoras, al mismo tiempo que puntualiza que las evaluaciones deben diseñarse de un modo adecuado desde el punto de vista cultural, en consonancia con las idiosincrasias de la región de Asia Sudoriental, habida cuenta en todo momento de su heterogeneidad (Parker y Fraillon, 2016).

Las competencias de expresión escrita abren las puertas a la medición de la creatividad, la curiosidad y la apreciación de la cultura

Cabe destacar que las LSLA no evalúan la expresión escrita, que representa una aptitud básica imprescindible para la comunicación, el aprendizaje futuro y la participación plena en la vida económica, política y social, así como en muchos aspectos de la vida diaria³³. En la era digital y en

33 Esto no se puede desvincular del hecho de que, tradicionalmente, las iniciativas de enseñanza y de alfabetización en masa se han centrado en la lectura en lugar de en la escritura.

el contexto de la economía del conocimiento, la comunicación personal y social se lleva a cabo cada vez con mayor frecuencia de manera escrita, por ejemplo, a través de teléfonos móviles y medios sociales. La evaluación de las aptitudes de expresión escrita o el uso de estas para medir ámbitos como la creatividad, la curiosidad y la valoración de la cultura también suele estar fuera del alcance de las LSLA. Esto se debe a que resulta especialmente difícil evaluar un texto escrito de un modo que siga posibilitando la comparación a nivel del sistema por causa de su carácter subjetivo.

Tras reconocer la importancia de las aptitudes de expresión escrita y el hecho de que se valoran en los exámenes, se están poniendo en marcha iniciativas encaminadas a incorporar esta dimensión a las LSLA. Pese a que se encuentra en fase de desarrollo, el programa SEA-PLM representa la primera iniciativa transnacional que mide las *competencias escritas*, entendidas como la capacidad de elaborar un mensaje con significado al generar textos escritos para expresarse y comunicarse con los demás, con el objeto de dar respuesta a las necesidades personales, sociales, económicas y cívicas. Esta definición reconoce que la expresión escrita va más allá de la capacidad de escribir palabras con una caligrafía legible. Esta concepción de la “expresión escrita” transmite la idea de que se lleva a cabo en un contexto, para un público y con un propósito. Engloba diversas aptitudes cognitivas, como generar y organizar ideas, aplicar vocabulario y hacer uso de los conocimientos relativos a las estructuras lingüísticas y las características textuales (SEA-PLM *et al.*, 2017, págs. 30 y 31).

Avanzando hacia la medición de la reflexión y las competencias colaborativas, así como de las aptitudes socioemocionales

En general, las LSLA no incluyen la medición de la reflexión y las competencias colaborativas. Sin embargo, son esenciales para la generación y la maduración de ideas (Hannon *et al.*, 2011), que son, cada vez con mayor frecuencia, el resultado de la cooperación entre uno o varios grupos de personas que reflexionan y trabajan juntas sobre la base de conocimientos existentes a fin de crear otros nuevos (OCDE, 2018a). Las competencias y la reflexión colaborativas (es decir, la capacidad de relacionarse con los demás, cooperar, trabajar en equipo, gestionar y solucionar conflictos, etc.) son difíciles de medir en las evaluaciones realizadas a nivel individual. Las personas jóvenes necesitan cada vez más este tipo de aptitudes para adaptarse a las demandas del mercado laboral, que evolucionan de manera vertiginosa.

Si bien la medición de la toma de decisiones, la coordinación a la producción en grupo por medio de evaluaciones individuales ha resultado ser especialmente complicada, se están llevando a cabo algunas iniciativas para reflejar las capacidades individuales en situaciones de colaboración. El proyecto PISA, por ejemplo, emplea pruebas informatizadas para evaluar las competencias colaborativas de los estudiantes a través de agentes simulados de manera digital que representan a miembros del equipo con diversas aptitudes y capacidades. Este enfoque consigue el alto grado de control y estandarización necesario para una correcta medición. Por otra parte, posibilita la observación de los estudiantes en diferentes situaciones de colaboración y permite realizar mediciones dentro las limitaciones de tiempo de la realización de la prueba (OCDE, 2017). No obstante, en líneas generales, estas iniciativas se encuentran todavía en gran

medida en una etapa incipiente, ya que existe un entendimiento aún limitado de las posibles dificultades.

Dado su valor intrínseco, las aptitudes socioemocionales constituyen otra esfera de medición a la que se está prestando más atención como factor determinante clave del progreso del aprendizaje. Por ejemplo, el LLECE está trabajando en estos momentos en un módulo socioemocional para la próxima evaluación regional, ERCE 2019. El módulo se probó de manera experimental en 2018 y tiene por objeto reflejar la concepción de la calidad de la educación promovido por la agenda Educación 2030 y estudiar el efecto de estas aptitudes en los resultados del aprendizaje, así como los factores determinantes de dichas aptitudes.

Aumento de la relevancia de las aptitudes digitales

El debate sobre la medición de las aptitudes digitales ha entrado solo recientemente en el terreno de la educación mundial. Remarcablemente, esto se debe, en gran parte, a la selección de las aptitudes digitales y las competencias relativas a los medios de comunicación y la información para la meta 4.4 de los ODS, que, a su vez, se centra en las aptitudes para el trabajo de las personas jóvenes y adultas. Esto explica por qué la labor conceptual y técnica que se ha llevado a cabo en el plano mundial ha estado en cierto modo desconectada de los debates sobre la evaluación de los niños en edad escolar.

Aunque está creciendo el número de iniciativas orientadas a evaluar las aptitudes digitales y de la TIC, su naturaleza es extremadamente heterogénea, tanto a efectos de su contenido como de las autoridades responsables (IEU, 2018c). Además, las experiencias relativas a las evaluaciones digitales parecen limitarse en gran medida a los países de ingresos altos, y escasean las iniciativas transnacionales centradas en los estudiantes de la educación básica. Como excepción, destaca el Estudio Internacional sobre Competencia Digital y Tratamiento de la Información (ICILS) de la IEA. No obstante, las demás evaluaciones vigentes suelen estar orientadas hacia la población adulta.

En cualquier caso, existe una tendencia común a estas iniciativas consistente en el énfasis en una cantidad relativamente limitada de aptitudes, como el uso de la información digital y las aptitudes técnicas, mientras que otras esferas, como la colaboración en línea y la resolución de problemas con la TIC, se abordan con menor frecuencia (Siddiq *et al.*, 2016). De nuevo, estas tendencias sugieren que, por su diseño, las LSLA solo pueden ayudar a evaluar un número relativamente limitado de aptitudes. Esto supone un gran problema en vista de que las aptitudes digitales, cada vez más complejas y sofisticadas, constituyen la base para seguir aprendiendo. En otras palabras, el aprendizaje en diversos ámbitos está condicionado en gran parte por la capacidad de cada persona de realizar tareas técnicas relacionadas con la TIC, pero también de solucionar problemas complejos, evaluar la información de forma crítica y otras actividades similares (UNESCO, 2018). A pesar de las dificultades de la medición, están surgiendo algunas iniciativas que tratan de reflejar con precisión un conjunto más exhaustivo de competencias ligadas a la TIC. Un ejemplo de ello es el proyecto Niños Digitales en Asia y el Pacífico (DKAP) de la Oficina de la UNESCO en Bangkok, que ha desarrollado hace poco el Marco de Competencias para la Ciudadanía Digital, un marco que abarca una amplia gama de ámbitos, como la seguridad digital, la inteligencia emocional y la creatividad. Se han realizado pruebas experimentales del

marco de evaluación en cuatro países y se espera que proporcione una comprensión sustentada empíricamente del uso de la TIC por parte de los niños, más allá de las aptitudes básicas³⁴.

Margen de maniobra: abarcando otras esferas de competencia

En resumen, como sugieren los ejemplos expuestos anteriormente, las organizaciones y los gobiernos nacionales están tratando de abarcar otras esferas de competencia, como las aptitudes transversales o transferibles, o las denominadas “habilidades del siglo XXI”, incluidas las habilidades “no cognitivas” o intersectoriales y las competencias relacionadas con el desarrollo sostenible, la ciudadanía en el mundo global y la salud, así como las vinculadas a la alfabetización digital. Puesto que se considera que dichas esferas engloban las principales aptitudes y competencias necesarias para ser trabajadores y ciudadanos efectivos en una sociedad basada en el conocimiento, la OCDE, la IEA y diferentes programas de evaluación regionales han puesto en marcha una serie de iniciativas orientadas a evaluarlas. Esto refleja la mayor aceptación de una concepción holística del aprendizaje en la que no solo se consideran cruciales y relevantes para el mundo actual los conocimientos generales básicos y las “aptitudes de base”, sino también competencias más específicas asociadas a la ciudadanía mundial, el desarrollo sostenible y la salud. Asimismo, favorece una serie de conocimientos, aptitudes, actitudes y valores globales necesarios para lograr una comunicación, una colaboración, una resolución de problemas y unas relaciones sociales efectivas.

Al mismo tiempo, es posible que los conocimientos y competencias en ámbitos relacionados con la socialización y la adquisición de valores, sean mejor capturadas mediante formas no tradicionales de evaluación incorporadas al proceso de enseñanza y aprendizaje, en tanto que se trata de competencias y conocimientos, que varían en función de cada cultura. Por consiguiente, a pesar del interés por abarcar dichas competencias, la conveniencia de adoptar medidas globales es cuestionable. Además, esta labor podría redundar en un aumento de la frecuencia de las pruebas y en un incremento considerable de los recursos financieros y el tiempo destinados a la evaluación.

La respuesta a los efectos de distracción o desviación de las LSLA no puede basarse exclusivamente en la realización de más pruebas. Más bien, es necesario buscar maneras de mejorar la integración de las LSLA en los sistemas de evaluación generales.

En suma, cabe reconocer que existe una tensión entre la necesidad de ampliar y diversificar el alcance de las LSLA y la imposibilidad e no-deseabilidad de evaluar cada ámbito de aprendizaje con ellas. De estas tensiones podría desprenderse que la respuesta a los efectos de distracción o desviación de las LSLA no puede basarse exclusivamente en la realización de más pruebas. Más bien, es necesario buscar maneras de mejorar la integración de las LSLA en los sistemas de evaluación generales, combinándolas con otras formas de evaluación para que la atención no se centre principal y exclusivamente en ellas.

34 Si desea obtener más información sobre esta iniciativa, consulte la página https://teams.unesco.org/ORG/fu/bangkok/public_events/Shared%20Documents/EISD/2017/Oct2017%20-%20KFIT%203%20Launch%20-%20Dig%20Citizenship/DKAP-project-brief.pdf

Restringiendo las personas que valoramos

En este apartado se ahonda en las limitaciones que presentan las LSLA en términos de inclusión. Tras explorar el potencial de las LSLA como instrumentos para la equidad, se analizan distintas maneras en que determinados grupos quedan excluidos de los ejercicios de evaluación o participan en ellos en condiciones comparativamente desfavorables. Por último, el apartado repasa diversas iniciativas llevadas a cabo recientemente con el fin de reforzar el carácter inclusivo de las evaluaciones, así como las dificultades y los riesgos que conllevan.

Evaluación al servicio de la equidad

La agenda Educación 2030 expone sin ambigüedades su compromiso con la equidad y la inclusión, consideradas las piedras angulares de una agenda verdaderamente transformadora. La equidad implica incluir a todas las personas en el proceso educativo y velar por que cada una de ellas, con independencia de sus orígenes, aprenda. Los llamamientos a elaborar una *agenda para la equidad del aprendizaje* (Wagner, 2018) constituyen un intento de reflejar la necesidad de aumentar los niveles de aprendizaje entre los sectores más desfavorecidos de la población.

Dicho planteamiento tiene grandes repercusiones en las iniciativas de medición. Se espera que el diseño de las evaluaciones tenga en cuenta las prioridades políticas en materia de equidad al inicio del proceso de evaluación y recabe información adecuada que ayude a vincular los resultados de las evaluaciones con las políticas orientadas a mitigar la desigualdad. En concreto, esta visión renovada de la educación requiere que los datos se desglosen, cuando sea posible, con arreglo a un amplio abanico de dimensiones (UNESCO, 2016c). Las LSLA pueden ayudarnos a comprender distintas dimensiones de la equidad en el sistema educativo, tales como la equidad entre regiones, la equidad entre escuelas, la igualdad de género, la equidad lingüística o la equidad asociada a la situación socioeconómica y a la pertenencia a un subgrupo concreto (p. ej., los estudiantes con discapacidad o las personas migrantes o refugiadas llegadas recientemente al país) (Brink, 2018).

Se espera que las evaluaciones sean inclusivas, que estén diseñadas de tal manera que sean pertinentes para el mayor número posible de miembros de la población a la que están dirigidas y de tal manera que ofrezcan una respuesta eficaz para todo el alumnado.

Por consiguiente, se espera que las evaluaciones sean inclusivas o, lo que es lo mismo, que estén diseñadas de tal manera que sean pertinentes para el mayor número posible de miembros de la población a la que están dirigidas y de tal manera que ofrezcan una respuesta eficaz para todo el alumnado (UNESCO, 2017b). En otras palabras, se espera que la muestra de alumnos no excluya intencionadamente a determinados grupos como, por ejemplo, las personas con discapacidad³⁵. Se espera igualmente que las personas que se sometan a las pruebas en distintas partes del país o en distintos países perciban las preguntas y las respondan de la misma manera, con independencia de su origen étnico, situación migratoria, género o lugar de residencia.

35 Véase Rowe, E. 2017. "Australia Country Case Study". Documento de antecedentes sobre la práctica del muestreo selectivo en Australia escrito para el *Informe de Seguimiento de la Educación en el Mundo 2017/2018*.

Garantizar estos principios de equidad e inclusión tiene importantes consecuencias para las iniciativas de medición. Por ejemplo, es necesario traducir y adaptar los materiales de las pruebas a fin de adecuarlos a los distintos grupos de población o a los niños con necesidades especiales. También hay que esforzarse más para que las poblaciones vulnerables o minoritarias tengan un mayor peso en los muestreos y velar por que el número de estudiantes pertenecientes a estos grupos cumpla con los requisitos mínimos para efectuar un análisis significativo. Asimismo, tal vez haya que adaptar los instrumentos y ajustar las condiciones de las pruebas de modo que los niños que pertenezcan a estos grupos gocen de igualdad de oportunidades a la hora de llevarlas a cabo. Del mismo modo, es preciso modificar los cuestionarios de antecedentes para que reflejen adecuadamente la información pertinente.

¿Quién queda excluido de las evaluaciones?

No siempre se cumplen las obligaciones en materia de inclusión. En realidad, el diseño de las evaluaciones y sus muestreos excluyen a muchos grupos vulnerables y marginados (estudiantes o escuelas concretos) por diversos motivos. Para empezar, las muestras no suelen incluir a los niños sin escolarizar (o a los niños matriculados en escuelas no registradas), ya que las LSLA se centran en el sistema educativo institucionalizado y solo las realizan niños matriculados en escuelas autorizadas y reconocidas oficialmente³⁶.

En ocasiones se excluye a escuelas enteras porque se encuentran en zonas remotas de difícil acceso o porque imparten educación a grupos de niños que están excluidos de la prueba (p. ej., escuelas para personas ciegas). Dentro de las escuelas también se excluye a niños concretos. Suelen verse excluidos los niños con discapacidades físicas permanentes moderadas o graves, los niños con discapacidades cognitivas, conductuales o emocionales, y los que carecen de suficiente experiencia con el idioma. Para agravar el problema, existe poca información y una comprensión limitada acerca de las medidas adoptadas por muchos programas de evaluación nacionales con el fin de combatir la exclusión y dar cabida a todos los alumnos. Aunque se trata de cuestiones cruciales, las orientaciones técnicas son escasas y a menudo se limitan a contextos de ingresos altos.

Las dinámicas de exclusión y representación inadecuada quedan especialmente patentes en países en los que existe una diversidad lingüística considerable. En tales contextos, el idioma de enseñanza constituye un reto y puede ser la clave para lograr que la educación sea más inclusiva para los grupos desfavorecidos (UNESCO, 2016b). Si bien resulta complicado determinar la proporción de estudiantes que reciben educación en su lengua materna (UNESCO, 2017d), una estimación sugiere que hasta el 40 % de los estudiantes no reciben educación en el idioma que hablan o entienden por completo (Walter y Benson, 2012). Las LSLA y otras modalidades de evaluación se realizan en general en el idioma de enseñanza, independientemente de si el niño lo entiende bien o no. Esto puede redundar en una exclusión involuntaria o en una menor participación de los grupos marginados. También es posible que los niños de estos grupos

36 Existe una conciencia cada vez mayor de la necesidad de incluir estas poblaciones en las LSLA. PISA para el Desarrollo, las EGRA y ciertas evaluaciones impulsadas por la ciudadanía son algunos de los últimos intentos que se han llevado a cabo con miras a incluir en ellas a los niños y jóvenes sin escolarizar (véase Yasunaga, 2014).

participen en las LSLA pero que sus niveles de aprendizaje reales permanezcan invisibles en el plano global y nacional (IEU, 2018a). Estas tendencias pueden ocultar las dificultades a que se enfrentan los países en relación con la equidad y la privación del derecho a la educación que sufren los grupos vulnerables. Por supuesto, las evaluaciones también ofrecen la posibilidad de llamar la atención sobre esas cuestiones. El bajo rendimiento escolar contribuye a que los funcionarios públicos y los profesionales del desarrollo tomen conciencia de que al menos las capacidades básicas de lectoescritura deben enseñarse en la lengua del alumno (Benson y Wong, 2015). La interpretación continúa siendo un factor clave para determinar el efecto decisivo de las LSLA y el potencial de estas a la hora de contribuir a un programa en materia de equidad.

Posibles riesgos de una adaptación excesiva

Recientemente se han tomado medidas orientadas a que las evaluaciones abarquen a los grupos marginados. El Informe Anual sobre la Situación de la Educación (ASER) en Pakistán ha llevado a cabo pruebas con niños con dificultades moderadas o graves para ver, oír, caminar, cuidarse, entender o recordar (Singal y Sabates, 2016). Por su parte, RTI International ha adaptado los instrumentos de las EGRA para que den cabida a los niños con deficiencias visuales y auditivas, y así mejorar la experiencia de estos en el aula. Desde 2015, el PISA ha tratado de ampliar el acceso a su evaluación. Sin embargo, como muestra el **Cuadro 5**, la intención de fomentar la equidad y la inclusión puede resultar problemática a su manera.

Cuadro 5. La adaptación de las pruebas: ¿un arma de doble filo?

En 2011, el consejo de administración del PISA manifestó su interés por buscar maneras de brindar acceso a dichas evaluaciones a los estudiantes con discapacidad intelectual, con discapacidad funcional (física) y con una experiencia insuficiente en el uso del idioma. Hasta entonces, las adaptaciones se habían limitado a grupos reducidos o a la realización de pruebas individuales. El PISA no ha permitido la aplicación generalizada de otros tipos de adaptaciones (como modificaciones gráficas, presentaciones orales, diccionarios bilingües, dictado de las respuestas, procesadores de texto, adaptaciones mediante la desnaturalización o herramientas adaptativas, etc.). Sin ellas, alumnos que de otro modo podrían realizar aportaciones sustanciales, deben quedar excluidos de la evaluación. No obstante, la adaptación que los estudiantes con necesidades educativas especiales solicitan con más frecuencia es la disposición de más tiempo o de un tiempo ilimitado, algo que no puede ofrecerse para rellenar los cuadernillos estándar de las pruebas del PISA.

Sin embargo, efectuar adaptaciones para incluir a grupos marginados o vulnerables en los procesos de muestreo conlleva ciertos riesgos. En primer lugar, existe el temor de que las necesidades de los estudiantes varíen tanto que resulte difícil garantizar la comparabilidad dentro de los países y entre ellos. Se teme también que pueda parecer que determinadas adaptaciones ofrecen a los estudiantes que hacen uso de ellas una ventaja comparativa sobre los que no las utilizan. Por último, las adaptaciones podrían crear un incentivo perverso para identificar a algunos alumnos como estudiantes con necesidades especiales, aunque no lo sean, simplemente para mejorar los resultados.

Fuente: Basado en OCDE, 2011.

Cada vez se adoptan más medidas encaminadas a traducir los instrumentos de evaluación y adecuarlos a contextos multilingües. Asimismo, todas las iniciativas transnacionales disponen de sofisticadas directrices de traducción y adaptación, y procuran, en la medida de lo posible, que las pruebas sean equiparables entre los distintos países e idiomas. Si bien en los países en desarrollo muy pocas LSLA se llevan a cabo en la mayoría de los idiomas de enseñanza o en todos ellos, se están realizando esfuerzos encomiables para incluir más idiomas en las

evaluaciones. Por ejemplo, en 2019, el PASEC efectuará las pruebas en los primeros niveles educativos en la lengua materna de los alumnos, mientras que, en el ciclo de 2013-14, solo las realizó en el principal idioma de enseñanza. Del mismo modo, Sudáfrica llevó a cabo su Evaluación Nacional Anual en otras 11 lenguas maternas hasta su supresión en 2014 (IEU, 2019a).

Efecto suelo

El denominado “efecto suelo” constituye otra limitación del diseño que puede disminuir la capacidad de las evaluaciones a la hora de respaldar un programa de equidad. El efecto umbral se produce cuando la mayoría de los participantes en una evaluación determinada obtienen resultados cercanos a la base de referencia. Como explican Wagner *et al.* (2018), hacer un hincapié excesivo en la comparabilidad a expensas de la validez local (según lo contemplado en las LSLA) puede provocar que las evaluaciones no proporcionen una comprensión adecuada de la situación de las poblaciones desfavorecidas, que tienen más probabilidades de obtener peores puntuaciones. Winthrop y Anderson (2013) formulan un argumento similar al señalar que, aunque las evaluaciones con un efecto suelo considerable puedan llamar la atención sobre ciertas cuestiones de interés, es probable que lo hagan a costa de su potencial analítico. En resumidas cuentas, la limitación de la validez local tiene un efecto especialmente negativo en la capacidad de las LSLA para respaldar la agenda de equidad.

Un hincapié excesivo en la comparabilidad a expensas de la validez local puede provocar que las evaluaciones no proporcionen una comprensión adecuada de la situación de las poblaciones desfavorecidas.

¿Los límites en el diseño acarrearán límites en el uso?

En términos generales, las tendencias descritas en el presente capítulo sugieren el carácter necesariamente parcial e imperfecto de los datos resultantes de las LSLA. Esto no debe entenderse como un defecto o una limitación exclusivos de las pruebas estandarizadas, ya que es probable que cualquier modalidad de reunión de datos priorice ciertas competencias sobre otras y se base, inevitablemente, en una definición selectiva de la calidad de la educación. No obstante, las tendencias expuestas invitan a ser cautelosos respecto de los riesgos que puede entrañar dar una credibilidad o una atención excesivas y descompensadas a los datos de las LSLA. Si no se combinan adecuadamente con otras fuentes de datos (cualitativas o cuantitativas), las limitaciones inherentes a las LSLA generan las condiciones propicias para hacer un uso indebido de ellas. Por tanto, puesto que los datos de las evaluaciones son, por definición, parciales e imperfectos, es probable que prestar una atención descompensada a las LSLA o no incorporarlas en otros componentes de los sistemas dificulte su capacidad de contribuir a un avance significativo hacia la mejora de la calidad y la equidad educativas.

4. Más allá de los datos como un fin en sí mismo: limitaciones en el uso de las evaluaciones de aprendizaje a gran escala

El capítulo anterior demostró que el diseño de las evaluaciones de aprendizaje a gran escala (LSLA) es la causa fundamental de algunas de sus consecuencias imprevistas. Sin embargo, la mayoría de las limitaciones asociadas a dichas herramientas tienen su origen en el uso o la falta de uso de los datos resultantes. Este capítulo comienza aclarando algunos conceptos ligados a los propósitos de las LSLA para después reflexionar acerca de los tres principales grupos de preocupaciones relativas al uso de los datos resultantes, a saber, su infrautilización, su sobreutilización y su combinación con medidas de rendición de cuentas (o su subordinación a ellas).

Usos de las evaluaciones de aprendizaje a gran escala: aclaración de términos y conceptos

Antes de analizar las consecuencias imprevistas de las LSLA es necesario efectuar un resumen general de los propósitos y los usos previstos de esas herramientas. Para ello, el presente apartado comienza aclarando ciertos conceptos respecto del tipo de información generado por las LSLA. A partir de esas precisiones, se examinan los múltiples usos y propósitos de las LSLA, así como las distintas partes interesadas que pueden beneficiarse de ellas.

Diferencia entre “resultados” y “evidencia”

Cuando se emplean los datos de las LSLA, un “resultado” es una conclusión derivada de un único análisis. Por ejemplo, las puntuaciones medias de niños y niñas en una materia o un ámbito de aprendizaje concretos o la distribución de un grupo particular en los distintos niveles de competencia. Por otro lado, la “evidencia” es un conjunto de conclusiones empleadas para responder a una política (o una cuestión práctica) con datos, normalmente describiendo tendencias, identificando patrones, señalando factores clave para explicar ciertas regularidades y poniendo de relieve posibles esferas de intervención. Toda respuesta a una política o una práctica debe basarse en evidencia o resultados de varios análisis o proceder de varias fuentes de datos y, lo que es más importante, debe contar con un componente de valoración y reflexión (Brink, 2018).

Por ejemplo, si la política en cuestión pretende garantizar la equidad entre niños y niñas y reducir las disparidades en su desempeño, la evidencia podrían incluir un análisis del modo en que dichas disparidades han aumentado o disminuido con el tiempo o de si existe una tendencia constante. Esta información, unida a un examen histórico de las medidas normativas emprendidas para afrontar el problema y a un estudio de los principales factores normativos delicados que afectan al desempeño, resultaría crucial para el diseño y la formulación de políticas (Brink, 2018). Por consiguiente, los resultados no hablan por sí solos y el asesoramiento sobre políticas basado en pruebas no se desprende *de forma natural* de los datos brutos.

Aclaración de la diferencia entre “análisis” y “uso”

Los datos obtenidos con las LSLA – en un único momento o en distintos ciclos de evaluación – proporcionan a los gobiernos pruebas con las que subsanar las ineficacias del sistema respondiendo a preguntas clave, como quién aprende qué y quién no, dónde, cuándo y por qué (Montoya, 2016). Los expertos e investigadores³⁷ emplean generalmente los resultados escolares y la información de los cuestionarios de antecedentes para describir los conocimientos y las aptitudes de una población específica. Eso suele implicar varios tipos de análisis. En primer lugar, es preciso entender los factores que influyen en los resultados escolares (p. ej., el contexto doméstico y escolar, así como las prácticas en dichos ámbitos) y si cambian con el tiempo. En segundo lugar, hay que identificar las tendencias generales de los resultados escolares y evaluar los progresos realizados en la consecución de metas específicas por medio de una serie de indicadores. En tercer lugar, es preciso poner de relieve las disparidades en las capacidades cognitivas de los distintos subgrupos de alumnos en función de la dimensión a que pertenecen, como factores socioeconómicos, cuestiones regionales, el género, la situación migratoria y la lengua materna.

Por su parte, los encargados de formular políticas utilizan los resultados o la evidencia de las LSLA con múltiples propósitos³⁸:

Seguimiento y evaluación – Hacer un seguimiento de una serie de cuestiones y tendencias con relación a la consecución de objetivos y metas nacionales, regionales y globales, y evaluar las políticas de educación.

Fundamento para la formulación de políticas – Orientar la formulación, el diseño y la implementación de políticas o mejorar la eficacia de las medidas normativas ya en vigor (véase la **tabla 5**).

Establecimiento de agendas – Establecer programas aumentando la sensibilización sobre determinadas cuestiones, fomentando el debate en el seno de la sociedad civil y en los círculos gubernamentales, identificando las esferas prioritarias de reforma y generando una “presión positiva” que incite a los gobiernos a emprender reformas o impulsar un cambio normativo, en especial cuando los resultados de las evaluaciones resulten sorprendentes o preocupantes.

37 Los analistas suelen pertenecer a organismos públicos, institutos públicos de evaluación, departamentos de universidades y grupos de investigación; organizaciones privadas especializadas; consorcios transnacionales de evaluación; y organizaciones internacionales. Su afiliación exacta, sin embargo, varía enormemente según las distintas tradiciones administrativas, educativas y culturales.

38 Una gran cantidad de académicos han estudiado los usos y objetivos de las LSLA. Para consultar un análisis de los elementos enumerados, véanse Tobin *et al.* (2015), Lingard y Grek (2007), Goldstein y Thomas (2008), Fullan (2010) y Kellaghan *et al.* (2011).

Rendición de cuentas – Hacer que una serie de agentes, incluidas las partes interesadas directamente en la educación (administradores, directores, docentes, juntas escolares o proveedores privados), y entes que podría decirse que son externos al sistema educativo (gobiernos, autoridades locales, etc.) rindan cuentas.

Análisis en profundidad – Añadir profundidad y perspectiva al análisis de los sistemas educativos y exigir a los encargados de la formulación de políticas y a los profesionales que lleven a cabo otros estudios específicos que reflejen de un modo más adecuado sus realidades locales y contextuales.

Si bien se sabe más o menos el uso que los encargados de la formulación de políticas, los expertos y los investigadores podrían hacer de los datos de las LSLA, existen otros grupos a los que también podrían interesarles los datos y conclusiones de las evaluaciones. Dichos grupos incluyen a docentes y directores; padres y alumnos; grupos interesados, como organizaciones no gubernamentales (ONG), especialistas en elaboración de planes de estudios, formadores de docentes o sindicatos de docentes; e investigadores y académicos que realicen análisis más minuciosos. La población general también puede aprovechar los resultados de estas evaluaciones siempre que se hayan adaptado a públicos sin conocimientos técnicos, por ejemplo, mediante la publicación de infografías sencillas en los medios sociales (para consultar un resumen de los posibles usuarios y públicos destinatarios, véase IEU, 2018c).

Tabla 5. Ejemplos de iniciativas políticas en las que pueden influir las evaluaciones de aprendizaje a gran escala, por nivel

SISTEMA	PLAN DE ESTUDIOS	ESCUELA	DOCENTES	HOGAR
<ul style="list-style-type: none"> • Asignación de recursos • Aplicación de reformas programáticas • Definición de objetivos en relación con los logros académicos 	<ul style="list-style-type: none"> • Elaboración del plan de estudios • Contenido y metodología del plan de estudios • Diseño del plan de estudios 	<ul style="list-style-type: none"> • Definición de las prioridades del cuerpo docente • Mejora de los servicios de apoyo al alumnado • Enriquecimiento del entorno escolar 	<ul style="list-style-type: none"> • Obtención de recursos para el desarrollo profesional o la mejora de las prácticas pedagógicas • Revisión de los cursos y las tareas 	<ul style="list-style-type: none"> • Conformación de programas destinados a fomentar la implicación de padres y madres

Fuente: Autores.

Cada uno de estos grupos puede emplear y aprovechar las LSLA de distintos modos. Por ejemplo, los docentes pueden basarse en los resultados de las evaluaciones para adaptar sus métodos de enseñanza y brindar un apoyo adecuado al alumnado, mientras que los padres y el público general pueden utilizarlos para hacer que las escuelas o las administraciones rindan cuentas (IEU, 2018c). Sin embargo, existen pocos conocimientos sobre cómo y en qué circunstancias es más probable que las distintas partes interesadas en la educación hagan un uso efectivo y adecuado de los datos disponibles, así como sobre sus consecuencias y los posibles riesgos y dificultades que pueden entrañar.

Infrautilización de los datos de las evaluaciones

Se asume que la evidencia resultante de las LSLA constituye un elemento clave para mejorar los sistemas educativos y el rendimiento de los estudiantes. No obstante, este cambio solo puede producirse si las principales partes interesadas hacen uso de la evidencia disponible, analizando los datos y comunicando los resultados a los principales grupos destinatarios, y utilizándolos para orientar las políticas públicas. En este apartado, abordamos factores que no se suelen tener en cuenta y que, a menudo, restringen el uso efectivo y suficiente de los datos, en particular, el restringido sentido de la apropiación respecto a la financiación, el diseño, la gestión y la difusión de las LSLA.

La reunión de datos no mejorará la calidad

Como ya se ha señalado, las LSLA se han convertido en componentes esenciales que proporcionan información fundamental sobre múltiples dimensiones de los sistemas educativos. De este modo, resulta necesario hacer un seguimiento de sus resultados, comprenderlos y entender cómo y por qué han cambiado, para adoptar o modificar las correspondientes políticas o estrategias con vistas a favorecer el aprendizaje. El proceso carecerá de efectividad si los datos resultantes no se analizan minuciosamente y se transmiten a las partes interesadas apropiadas para que puedan usarlos a fin de adoptar nuevas medidas al respecto (Kellaghan y Greaney, 2001).

Sin embargo, la utilización y difusión de los resultados no pueden darse por sentadas (Tobin *et al.*, 2015; IEU, 2018c). Por lo general, el proceso de evaluación del aprendizaje culmina con la redacción de un informe que resume las conclusiones y que quizá se presenta en un acto público, pero lo habitual es que los resultados dejen de comentarse después de eso. Con demasiada frecuencia, los gobiernos, las organizaciones regionales o internacionales y otros usuarios continúan centrándose en explicar el estado actual de la educación y las tendencias o en presentar información relativa únicamente a una serie de indicadores predeterminados, en lugar de buscar la mejor forma de emplear las conclusiones para mejorar el contexto y el proceso del aprendizaje. Otro obstáculo para la utilización productiva de los resultados es que los docentes y educadores no siempre consideran que las LSLA sean útiles, por lo que estas pueden acabar teniendo una escasa repercusión en los procesos de aprendizaje y enseñanza. Si bien la presentación de informes es un paso necesario (pues requiere comprender los resultados y las complejas relaciones existentes entre el aprendizaje y las variables que influyen en él), la obtención de datos no conduce de forma automática a la mejora del aprendizaje. Cuando las evaluaciones no se emplean para impulsar un cambio, la medición de los resultados del aprendizaje corre el riesgo de convertirse en un fin en sí mismo.

La utilización y difusión de los resultados no pueden darse por sentadas. Si bien la presentación de informes es un paso necesario, la obtención de datos no conduce de forma automática a la mejora del aprendizaje.

Como se ha indicado anteriormente, para la elaboración y realización de las LSLA se requieren ingentes recursos, pues se precisa una financiación, una infraestructura y una preparación considerables. El costo total puede ser considerable para los países de bajos ingresos y de

ingresos medianos bajos³⁹. Por tanto, resulta crucial aprovechar los datos resultantes para maximizar la utilidad de la inversión y garantizar que la evaluación no sea percibida como un derroche que supone una carga innecesaria para las arcas públicas⁴⁰.

Reconociendo los obstáculos para facilitar un uso efectivo

Si bien se espera que las LSLA cumplan diversos propósitos, poco se sabe acerca del modo y las condiciones en que sus resultados se pueden usar con eficacia. Algunos de estos factores determinantes se han examinado en detalle, y las preocupaciones al respecto pueden consultarse en obras influyentes sobre las LSLA, como la publicación *Using assessment to improve the quality of education* de Kellaghan y Greaney (2001) (véase también IEU, 2018c, y Raudonyte, 2019, para obtener una explicación más actual de las barreras que impiden la utilización de los datos de las LSLA). Las limitaciones que se suelen reconocer son muy diversas. Entre ellas figuran dificultades financieras y técnicas (p. ej. los conocimientos limitados o fragmentados de las partes interesadas en la materia o la mala calidad de las evaluaciones en vigor); la ausencia de implicación o voluntad políticas; la integración insuficiente en los procesos de formulación de políticas públicas; la escasa armonización entre las evaluaciones y otros componentes del sistema educativo; la exigua fiabilidad o relevancia de la información; la superficialidad de los análisis; el uso de formatos de presentación inapropiados, y la difusión escasa o insuficiente (IEU, 2018a, 2018c; Lockheed, 2015; Tobin *et al.*, 2015).

Al mismo tiempo, se teoriza o reflexiona muy poco sobre otras limitaciones y, lo que es más importante, se sigue hablando de las múltiples barreras que dificultan el uso sin apenas ponerlas en contexto. En consecuencia, existe el riesgo de pasar por alto las dinámicas concretas que afectan a los países de bajos ingresos o que han adoptado las LSLA más tarde. En este apartado, se expone en detalle un grupo concreto de limitaciones que afectan específicamente a los países en desarrollo: el restringido sentido de la apropiación (*country ownership*). El uso restringido de los resultados de las evaluaciones no puede atribuirse exclusivamente a la ausencia de voluntad política. La falta de *oportunidades, apoyo, recursos o capacidad* de las principales partes interesadas en la educación para acceder a los datos disponibles, interpretarlos y actuar en consecuencia tiene

El uso restringido de los resultados de las evaluaciones no puede atribuirse exclusivamente a la ausencia de voluntad política. La falta de oportunidades, apoyo, recursos o capacidad de las principales partes interesadas en la educación para acceder a los datos disponibles, interpretarlos y actuar en consecuencia tiene un gran peso.

39 La dotación de personal, la capacitación y el desarrollo de la capacidad, los viajes y los equipos, la presentación de informes y la difusión de la información son algunos de los costos financieros evidentes (Wagner, 2011; Lockheed, 2015). Además, para formar parte de algunas iniciativas transnacionales, es preciso abonar una comisión de participación. El costo directo total oscila entre los 400.000 y los 2 millones de dólares de los Estados Unidos (Clarke, 2016).

40 Se ha debatido acerca de si los beneficios de las evaluaciones valen lo que cuestan. Hay quien ha afirmado que se trata de una de las innovaciones de la reforma educativa menos caras (Clarke, 2012) o que su coste es relativamente reducido en comparación con el coste de ejecutar políticas o de tomar decisiones desinformadas o equivocadas (IEU, 2016a). No obstante, otros académicos sostienen que todo análisis de la relación coste-beneficio de dichos ejercicios entraña cierto grado de valoración normativa y que el discurso público no ha tratado suficientemente esas cuestiones (Engel y Rutkowski, 2018).

un gran peso. Así ocurre en particular en los países de menores ingresos, donde la capacidad para integrar las iniciativas nacionales y transnacionales en los sistemas educativos a menudo va a la zaga a causa de la escasa institucionalización (IEU, 2018c).

Consecuencias de un sentido de la apropiación restringido

En muchos países, la falta de armonización entre la producción de datos y su uso es una consecuencia directa del restringido sentido de la propiedad respecto a los datos de las evaluaciones. Se puede decir que, si los países no perciben que tienen algún tipo de influencia sobre el proceso de evaluación, serán menos proclives a usar los datos y a actuar basándose en las pruebas. Esta falta de sentido de la apropiación es, a su vez, el resultado de la limitada participación e influencia de los países en las diferentes partes del proceso de evaluación, en particular en la financiación, el diseño, la gestión de datos y la difusión de los resultados.

Evaluaciones con financiación externa. En las evaluaciones de aprendizaje nacionales y transnacionales, y, especialmente, en los países en desarrollo, se observa una tendencia general hacia las evaluaciones de aprendizaje con *financiación externa*. Por ejemplo, al menos dos terceras partes de las evaluaciones de aprendizaje nacionales están financiadas con el apoyo (en parte o en su totalidad) de donantes extranjeros o externos (IEU, 2019a), como la Alianza Mundial para la Educación, el Fondo de las Naciones Unidas para la Infancia (UNICEF), la USAID y el Banco Mundial.⁴¹ En lo que respecta al PASEC, desde 2014 la CONFEMEN sufraga la mitad de los gastos de las evaluaciones transnacionales, mientras que el gobierno deberá financiar el resto para participar. Las repercusiones de las evaluaciones con financiación externa podrían explicar el carácter irregular y esporádico de las evaluaciones en muchos países. Un aspecto todavía más preocupante es el hecho de que las evaluaciones que dependen de financiación externa pueden contribuir a la falta de armonización entre las evaluaciones nacionales y otros componentes del sistema educativo, lo que propicia el restringido sentido de la apropiación respecto a los resultados de las evaluaciones.

Proceso de diseño y gestión de las evaluaciones controlado. Cuando las evaluaciones se externalizan a ONG u organismos internacionales, los países asumen muy poca responsabilidad y apenas controlan el diseño de los instrumentos de evaluación y la gestión de los datos resultantes, en especial porque los representantes y expertos del país prácticamente no contribuyen al desarrollo de dichos instrumentos. Tal es el caso de las iniciativas de evaluación transnacionales en las que el organismo líder, junto con un contratista dedicado a los datos, estandarizan los conjuntos de datos de los países participantes antes de facilitárselos a los equipos nacionales para que efectúen sus análisis. Aunque así se garantiza la estandarización a la hora de tratar los datos, podría percibirse que las organizaciones gestoras “controlan” y “poseen” los datos.

Calendarios de difusión estrictos. También es habitual en las iniciativas transnacionales que los países tengan poco que decir sobre los datos o análisis que se difunden sobre ellos y sobre los

41 En América del Norte y Europa Occidental, ninguna evaluación cuenta con fondos externos. Un pequeño porcentaje de países latinoamericanos reciben apoyo externo para sus evaluaciones (11 %), pero esta cifra es más elevada en otras regiones: Asia y el Pacífico (30 %), Europa Oriental (30 %), África Subsahariana (36 %) y los Estados Árabes (50 %) (Clarke, 2017).

niveles de aprendizaje de la población destinataria. Es posible que solo se les permita difundir sus informes nacionales una vez que se haya publicado el informe comparativo internacional (OCDE, 2009; IEA, 2019). Esto se debe a los acuerdos que establecen habitualmente los gobiernos y los organismos para llevar a cabo iniciativas transnacionales, los cuales otorgan a los organismos el poder de difundir los resultados y sus interpretaciones (con independencia de si los países están contentos con ellas) con miras a evitar un conflicto de intereses⁴².

El escaso sentido de la apropiación sobre los datos no solo puede acarrear su infrautilización y la inacción política, sino que también puede llevar a los gobiernos, a las autoridades educativas y a otros agentes a interpretar o emplear estas evaluaciones de forma errónea.

El sentido de la apropiación de los datos resulta menos problemático en las evaluaciones nacionales, ya que normalmente los países controlan la estrategia, el diseño y el análisis de la evaluación. No obstante, los donantes que financian el programa de evaluación, especialmente en los países de bajos ingresos e ingresos medianos, pueden encargarse incluso del diseño y el desarrollo de los instrumentos de prueba o influir enormemente en ellos. Existe evidencia de que, en algunos países, los resultados de las evaluaciones nacionales no se validan ni se difunden adecuadamente, lo que disminuye su valor político (Unión Europea, 2014). Independientemente del tipo de evaluación, cabe señalar que el escaso sentido de la apropiación sobre los datos no solo puede acarrear su infrautilización y la inacción política, sino que también puede llevar a los gobiernos, a las autoridades educativas y a otros agentes a interpretar o emplear estas evaluaciones de forma errónea. La ausencia de un sentido de propiedad, por tanto, puede dar lugar a equívocos, ya que puede distorsionar las prioridades políticas y sugerir un rumbo inadecuado para la acción política.

Sobreutilización de los datos de las evaluaciones

En el presente apartado, examinamos los riesgos derivados del uso y la interpretación excesivos de los datos de las evaluaciones, es decir, los que surgen cuando los resultados de estas suscitan un interés, una atención o una credibilidad desmesurados. Reflexionamos sobre el potencial que tienen las dinámicas de uso excesivo para desorientar, desviar o distraer las decisiones y conductas de un amplio abanico de partes interesadas y usuarios de las LSLA y, en especial, de los encargados de la formulación de políticas, los docentes y los investigadores. En este apartado se exponen, en primer lugar, los distintos formatos, canales y agentes dedicados a la difusión de datos. A continuación, se examinan los cuatro mecanismos con los que los datos de las evaluaciones fomentan un uso excesivo. Estos consisten, en concreto, en incentivar acciones políticas orientadas a manipular las cifras, promover el préstamo no fundamentado de políticas, provocar la identificación de relaciones de causalidad engañosas, y fomentar el uso de planes de financiación basados en los resultados.

42 Aunque los países tal vez tengan la oportunidad de eliminar algunas variables antes de la difusión o publicación de los datos, solo podrán hacerlo si logran demostrar que dichos datos ponen en peligro el anonimato de un alumno, una escuela o un docente, o que se han producido errores técnicos que los contratistas encargados de los datos no pueden resolver (OCDE, 2018b, pág. 1).

Múltiples formas de difusión

Una vez analizados los datos de las LSLA, los resultados o conclusiones se resumen en uno o más productos concebidos para su difusión, tales como informes analíticos, documentos de políticas, informes y comunicados de prensa, bases de datos de evaluación e incluso publicaciones en blogs y medios sociales destinados a personas sin conocimientos técnicos. Posteriormente, estos productos se difunden de diversas maneras. Por supuesto, la divulgación de los resultados de las evaluaciones no necesita de *todos* estos productos.

En el caso de las evaluaciones transnacionales del aprendizaje, la difusión y comunicación de la información se lleva a cabo de una forma sistemática y relativamente homogénea. La divulgación de las conclusiones queda prohibida hasta la fecha de publicación de los resultados, fijada previamente. Después, se publican periódicamente nuevas conclusiones e informes temáticos, se da acceso público a la base de datos y se publican artículos en periódicos nacionales y blogs internacionales. Por el contrario, los resultados de las evaluaciones nacionales se divulgan de una forma más heterogénea. Solo una de cada cuatro evaluaciones se incorporará a una base de datos pública, mientras que solo una de cada tres aparecerá en los medios de comunicación (IEU, 2019a).

El papel de los medios de comunicación

Independientemente de cómo decidan divulgar los resultados las autoridades implicadas, la influencia de las LSLA también depende de la repercusión que tengan en la opinión pública (Feuer, 2012; Addey, 2018; Yasukawa *et al.*, 2017). Esta depende, en gran medida, de la atención que los medios de comunicación presten a las LSLA. En este caso, los medios de comunicación engloban la prensa tradicional y los medios de radiodifusión, así como los medios digitales e interactivos. La elevada repercusión y exposición en los medios permite a las LSLA influir en los elementos que el público considera valiosos en la educación (Hamilton, 2017; véanse ejemplos de tales dinámicas en Dixon *et al.*, 2013; Steiner-Khamsi *et al.*, 2017, y Waldow *et al.*, 2014). En consecuencia, el discurso dominante de las evaluaciones transnacionales del aprendizaje más generalizadas y difundidas (es decir, el PISA, el TIMSS y el PIRLS) determina qué se valora (y se financia) en un sistema educativo. Si bien estas dinámicas no son problemáticas en sí mismas, pueden reforzar una dependencia excesiva de los datos de las evaluaciones y desviar la atención de otras cuestiones relevantes, lo que podría dar lugar a inferencias y identificación de relaciones de causalidad engañosas. No obstante, es preciso ahondar en el asunto para comprender cómo los distintos medios de comunicación contribuyen a la construcción del discurso público en torno a las LSLA (Hamilton, 2017), en especial porque estos medios presentan diferencias notables en términos del tipo de público al que se dirigen, las personas que participan en ellos, la precisión de la información que presentan y el uso del lenguaje emocional que hacen, entre otros aspectos.

La elevada repercusión y exposición en los medios permite a las LSLA influir en los elementos que el público considera valiosos en la educación.

Incentivando el cambio política centrándose en las cifras

Como ya se ha observado, se espera que las LSLA sirvan de fundamento para las prioridades y los programas de los encargados de la formulación de políticas y, de este modo, contribuyan al debate relativo a las políticas educativas. Sin embargo, esta influencia en la formulación de políticas conlleva grandes riesgos si la acción política persigue el único objetivo de repercutir en las cifras y mejorarlas, en lugar de lograr un cambio sustancial a nivel estructural.

Es más probable que se produzcan estas dinámicas en el marco de las evaluaciones transnacionales debido, en gran medida, al formato de cuadro de clasificación (*ranking*) con el que, con frecuencia, se publican los resultados de estos ejercicios. Estos cuadros se han convertido en un mecanismo habitual de los medios de comunicación para comparar las condiciones y los desempeños educativos, ilustrar los avances realizados y poner de manifiesto los obstáculos existentes entre los distintos países y dentro de ellos. Esta forma de presentar los resultados —colocando los países con mejor desempeño arriba y aquellos con peores resultados, abajo— permite a los encargados de la formulación de políticas comparar el rendimiento de sus niños en edad escolar con los de otros países. Por tanto, los cuadros de clasificación sirven como indicador de los progresos nacionales. El objetivo de Australia de figurar entre los primeros cinco países del PISA en 2025, el propósito de Tailandia de alcanzar el nivel de rendimiento promedio de la OCDE en el PISA de aquí a 2021 (Breakspear, 2014) y el objetivo de los Estados del Golfo de mejorar su posición en la clasificación del PISA (UNESCO, 2017f) son algunos ejemplos ilustrativos del énfasis en mejorar las cifras.

Al mostrar qué países encabezan el cuadro y cuáles están a la zaga, este tipo de datos comparativos puede incrementar la presión ejercida sobre los gobiernos para mejorar el rendimiento escolar en las distintas dimensiones asociadas a los resultados del aprendizaje, incluso en los países con peor calificación. Por consiguiente, las comparaciones transnacionales pueden alentar la puesta en marcha de iniciativas políticas encaminadas a mejorar los resultados. Existen múltiples ejemplos empíricos de tales dinámicas. Por ejemplo, la publicación de los resultados del TIMSS (1999) sirvió para presionar a los Estados Unidos y el Reino Unido para que efectuaran reformas (Gorard, 2001), y los del PISA (2000) pusieron de relieve la existencia de grandes disparidades en Alemania, lo que desencadenó un importante debate público acerca de la necesidad de emprender reformas.

Si bien esas tendencias no son inherentemente problemáticas, los análisis suelen centrarse de manera excesiva en las clasificaciones y las puntuaciones medias, en lugar de en los factores que explican las diferencias de desempeño entre los distintos países (UNESCO, 2017a; Rutkowski y Rutkowski, 2016; Torney-Purta y Amadeo, 2013). El enfoque centrado en las clasificaciones y puntuaciones puede desorientar o desviar las iniciativas. Este es el caso, en particular, de los cambios políticos provocados por las dinámicas de préstamo (mal planteado) de políticas, las cuales se basan en un entendimiento muy superficial y descontextualizado de las repercusiones que tienen las intervenciones normativas y de su capacidad para influir en las puntuaciones de las pruebas.

Promoviendo el préstamo de políticas

Cada vez hay más indicios de que, en los últimos años, algunos países han implantado reformas educativas o llevado a cabo iniciativas políticas basándose en su desempeño en las evaluaciones (Fischman y Topper, 2017; Baird *et al.*, 2016)⁴³. Los cuadros de clasificación animan a los países a fijarse metas en función de los objetivos conmensurables alcanzados por otros sistemas educativos, por lo que es probable que los impulsen a imitar las políticas y prácticas aplicadas en otros lugares o a aprender de ellas (OCDE, 2016a). En este sentido, las listas y los cuadros de clasificación se han convertido en importantes instrumentos políticos para acelerar el cambio y la innovación (Steiner-Khamsi, 2003). Por consiguiente, es posible que las LSLA no solo incentiven los cambios políticos y que fomenten también prácticas de préstamo de políticas, incitando a los países a adoptar las prácticas y políticas de los que se encuentran en la parte

Los cuadros de clasificación animan a los países a fijarse metas en función de los objetivos conmensurables alcanzados por otros sistemas educativos, por lo que es probable que los impulsen a imitar las políticas y prácticas aplicadas en otros lugares o a aprender de ellas.

alta de las clasificaciones, como Finlandia o Shanghái, y generando un “peregrinaje educativo” en el que estos lugares se convierten en un destino de viaje popular entre los encargados de la formulación de políticas educativas, investigadores, docentes u otros agentes de los sistemas educativos (Sahlberg, 2007; Addey, 2014).

Como consecuencia de esta tendencia, suscita preocupación el hecho de que, aunque las evaluaciones sean cruciales para impulsar el desarrollo educativo, existe el riesgo de que los sistemas educativos y sus asociados se basen excesivamente en las pruebas para llevar adelante sus reformas. Algunos investigadores han documentado cómo estas dinámicas pueden tener

un efecto distorsionador o de distracción entre círculos políticos. En última instancia, pueden llevar a los encargados de la formulación de políticas a elegir normativas concretas de países con un desempeño óptimo sin tener en cuenta su contexto cultural o la coherencia general de las políticas (Breakspear, 2014), o a adoptar políticas y soluciones rápidas atractivas desde el punto de vista político (Lingard y Sellar, 2013), en lugar de promover un aprendizaje político significativo y sostenible que tenga debidamente en cuenta los propósitos de la escolarización. Por ejemplo, Gorur y Wu (2014) han documentado cómo la ambición de figurar entre los primeros cinco países del PISA ha llevado a Australia a intentar imitar a los países de Asia Oriental con mejores resultados, a pesar de no disponer apenas de información sobre las particularidades de dichos sistemas, de los que prácticamente solo se conocen sus puntuaciones medias y su clasificación.

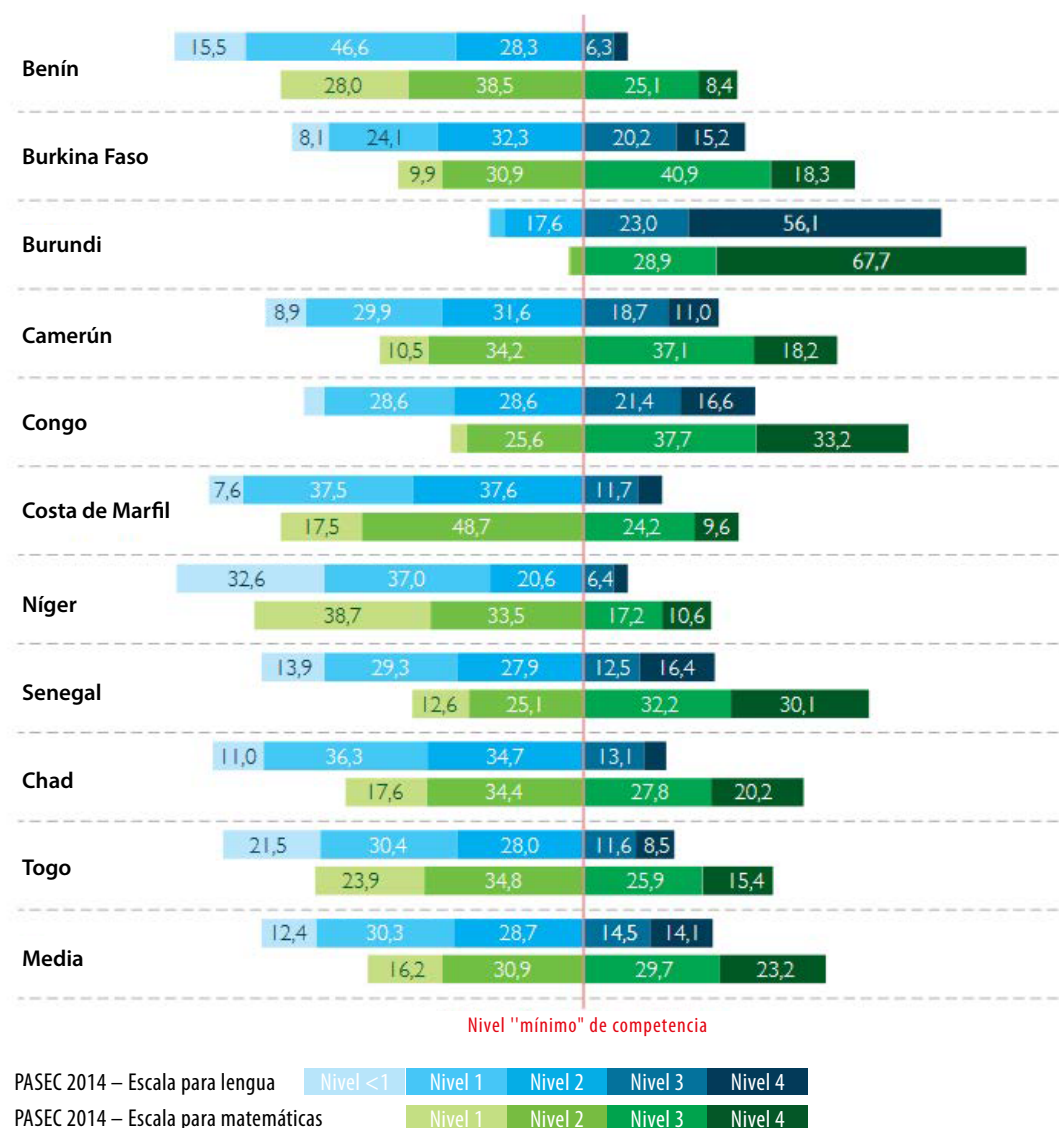
43 También hay cada vez más evidencia de que los gobiernos se sirven de su participación en iniciativas transnacionales para legitimar o acelerar sus reformas políticas, incluidos programas políticos preestablecidos y con motivación ideológica que de otra manera resultaría difícil justificar en términos puramente técnicos (véase Raudonyte, 2019).

Estas dinámicas son agravadas por las limitaciones y desigualdades que se dan en el uso de *big data* relativa a la educación para la elaboración de políticas basada en la evidencia, debidas en parte a la insuficiente capacidad institucional de algunos países para analizar dichos datos y vincular los resultados a las políticas. En consecuencia, estos países tal vez tomen como fundamento las conclusiones de los estudios de otros países y regiones, aun cuando es posible que no sean aplicables a sus sistemas educativos (UNESCO, 2017e). En este sentido, resulta alentador que determinados organismos, como la CONFEMEN, hayan empezado hace poco a presentar los resultados de los países en orden alfabético, en lugar de ordenándolos en función del rendimiento. Este pequeño cambio puede ayudar a neutralizar las dinámicas expuestas. En la **Figura 2** se muestra un ejemplo de presentación de resultados relativos al rendimiento académico por orden alfabético.

Investigación y causalidad

El interés de la comunidad académica en las LSLA ha ido en aumento, como demuestra la creciente cantidad de comentarios basados en opiniones y artículos académicos revisados por pares que se publican acerca de ellas. Las publicaciones, de diversa índole, se han centrado en las implicaciones de las LSLA, el uso de datos en las políticas nacionales y otros análisis posteriores basados en estudios empíricos (Fischman y Topper, 2017).

Un ámbito que recientemente ha ganado visibilidad en el debate relativo a las políticas educativas internacionales es el estudio del efecto que tiene la calidad de la educación (calculada a partir de las puntuaciones medias obtenidas en las pruebas de las evaluaciones de aprendizaje transnacionales) sobre el crecimiento económico. El tema es objeto de acalorados debates entre los académicos y las organizaciones internacionales. Por ejemplo, Hanushek y Woessmann (2015) han sostenido que la mejora de las aptitudes de una población (medidas mediante la mejora de los resultados del aprendizaje de las evaluaciones transnacionales entre un ciclo y el siguiente) redundaría en un crecimiento considerable del PIB. Este podría ser un mensaje atractivo con grandes repercusiones que podría llevar a creer que la mejora de las puntuaciones obtenidas en las pruebas se traducirá en beneficios económicos. Otros analistas, sin embargo, han señalado que dicho mensaje puede considerarse demasiado simplista por dos motivos. En primer lugar, el aprendizaje no termina en la escuela, sino que es un proceso ubicuo que evoluciona constantemente a lo largo de la vida (OCDE, 2010). En segundo lugar, los resultados económicos de un país dependen de una serie de variables que trascienden el ámbito del rendimiento académico y los logros educativos (Feuer, 2012). Por lo tanto, la afirmación de que existe una relación causal entre el rendimiento académico y el desempeño económico de un país puede resultar engañosa. Investigadores como Komatsu y Rappleye (2017) muestran que la correlación entre ambos factores es débil y sugieren que prácticamente no existe una relación causal entre ellos. Del mismo modo, Feuer (2012) ha demostrado en el pasado que las aseveraciones de que existe una relación entre los resultados académicos y los resultados económicos de los países no cuentan con suficiente apoyo empírico. Es más, señalan que las puntuaciones de las evaluaciones no constituyen fuentes de datos válidas para estimar el bienestar económico general de una nación.

Figura 2. Porcentaje de alumnado por nivel de competencias lingüísticas y matemáticas alcanzado en los primeros cursos de la enseñanza primaria

Fuente: PASEC, 2015.

Uno de los principales obstáculos para investigar la capacidad de predicción de las LSLA reside en la dificultad de ajustar por diversos factores de confusión. Por ejemplo, puede considerarse que los resultados conseguidos en las LSLA constituyen un buen modo de prever la matriculación en la enseñanza superior y la finalización de esta; como sugieren las pruebas extraídas de encuestas longitudinales llevadas a cabo en distintos países, el resultado obtenido en el PISA es un indicador válido y sólido de las posteriores experiencias educativas (OCDE, 2010). No obstante, también podría suceder que un mejor desempeño económico

se traduzca en un aumento de la escolarización y en la mejora de los resultados académicos. O quizás el hecho de que haya más personas con un mayor nivel educativo redundaría en un mejor desempeño económico (Earle, 2010).

En resumidas cuentas, existen razones de peso para ser más cautos al hablar de causalidad. Además, como destacaron Komatsu y Rappleye (2017), resulta necesario un debate global más riguroso sobre las políticas educativas, y valorar más detenidamente cómo se emplean las estadísticas y los datos en la formulación de políticas. Asimismo, la importancia relativa de los datos de las evaluaciones en comparación con otras fuentes de información es otro aspecto sobre el que hay que seguir reflexionando, a fin de que los estudios se basen en un conjunto diverso de evidencia.

Financiación basada en los resultados

La comunidad para el desarrollo también puede utilizar de manera excesiva las evaluaciones del aprendizaje, en especial cuando la calidad y los resultados de las LSLA determinan el nivel de financiación que los donantes externos aportan a los países. Por ejemplo, el Plan Estratégico de la Alianza Mundial para la Educación para 2016-2020, cuya prioridad es mejorar el aprendizaje y la equidad reforzando los sistemas educativos, fomenta el desarrollo de los sistemas nacionales de evaluación. La Alianza Mundial para la Educación pone énfasis en los datos y resultados, y utiliza un marco basado en resultados para medir los avances y exigir a todos los socios de la alianza que rindan cuentas (Alianza Mundial para la Educación, 2016). De esta manera, la Alianza podría dejar de proporcionar financiación si no se realizasen evaluaciones o no se demostrase una mejora en el desempeño de los alumnos. Aunque este tipo de planes tienen cierto potencial en lo que respecta a la promoción de los cambios normativos, también conllevan importantes riesgos. En particular, los planes de financiación basados en resultados pueden dar lugar a prácticas orientadas a “burlar el sistema” (UNESCO, 2017d), tales como realizar un muestreo selectivo, manipular la selección de los grupos que participarán en las pruebas o seleccionar a los estudiantes con mayor rendimiento para que se sometan a la prueba y así asegurar la financiación.

Los planes de financiación basados en resultados pueden dar lugar a prácticas orientadas a “burlar el sistema”, tales como realizar un muestreo selectivo, o manipular la selección de los grupos que participarán en las pruebas.

Asociación de las evaluaciones a la rendición de cuentas

Varios de los riesgos que se achacan habitualmente a las LSLA son, en realidad, consecuencia de combinarlas con marcos de rendición de cuentas con grandes repercusiones (o de supeditarlas a ellos). En este apartado, se estudia la naturaleza de los mecanismos que combinan estos instrumentos políticos con los principales riesgos y retos asociados a tales programas. El apartado concluye con una advertencia sobre la importancia relativa de los marcos globales de rendición de cuentas con grandes repercusiones (*high-stakes*). A continuación, se reflexiona sobre la necesidad de comprender mejor las demás variedades de LSLA, principalmente los programas basados en muestras y con escasas repercusiones (*low-stakes*).

Las LSLA como componente de las políticas de rendición de cuentas

Por lo general, se considera que las LSLA tienen pocas repercusiones para quienes se someten a ellas. Esto las diferencia de los exámenes nacionales, que se emplean con fines de selección, clasificación o certificación al final de un ciclo educativo (IEU, 2019b). Sin embargo, aunque las LSLA apenas tengan repercusiones para el alumnado, pueden tener consecuencias para otras partes interesadas, como los docentes, los directores, las escuelas o las administraciones educativas.

Así ocurre en los denominados marcos de rendición de cuentas basados en pruebas. Este tipo concreto de rendición de cuentas se caracteriza por fundamentarse en los datos obtenidos en pruebas estandarizadas y por dar una clara prioridad a los resultados que, por lo general, son los resultados del aprendizaje. En los marcos de rendición de cuentas basados en pruebas, algunas partes interesadas —en particular los docentes, los directores y las escuelas— deben rendir cuentas por del desempeño de sus alumnos con respecto a unos niveles predefinidos, el cual se mide en función de los resultados que obtengan en las LSLA (Verger y Parcerisa, 2017; Anderson, 2005; Hamilton *et al.*, 2002). Las partes interesadas pueden ser premiadas cuando los resultados sean buenos y penalizadas o sancionadas cuando estos se consideren insatisfactorios. Las consecuencias derivadas de estos resultados son muy dispares. La incapacidad de alcanzar un nivel mínimo aceptable puede traducirse en el cierre de una escuela, un cambio en su gestión y el despido o el traslado de docentes y administradores. En cambio, el alto rendimiento se premia con diversos incentivos materiales u organizativos, como bonificaciones monetarias, oportunidades de ascenso o mayores niveles de autonomía escolar.

Incentivos perversos cuando las puntuaciones en las pruebas se convierten en el aspecto central

Las implicaciones y las consecuencias imprevistas que tienen las disposiciones de rendición de cuentas asociadas a las evaluaciones del aprendizaje se han examinado en detalle en publicaciones especializadas, en particular en lo que se refiere a los planes con grandes repercusiones en los que los resultados de las evaluaciones tienen consecuencias claras y directas para los docentes, los directores y las escuelas.

Como se ha sugerido anteriormente, si se aplican sanciones y se ofrecen recompensas de forma explícita, las puntuaciones en las pruebas pueden convertirse en el eje central de la escolarización hasta el punto de que el tiempo de docencia se centre en practicar para la prueba en lugar de en el contenido marcado por los estándares curriculares (Weiss *et al.*, 2002). Esto puede constituir además un incentivo perverso para las escuelas y los docentes, que pueden adoptar prácticas cuestionables para obtener las mejores puntuaciones posibles. Por ejemplo, las escuelas pueden seleccionar

Si se aplican sanciones y se ofrecen recompensas de forma explícita, las puntuaciones en las pruebas pueden convertirse en el eje central de la escolarización hasta el punto de que el tiempo de docencia se centre en practicar para la prueba en lugar de en el contenido marcado por los estándares curriculares. Esto puede constituir además un incentivo perverso para las escuelas y los docentes, que pueden adoptar prácticas cuestionables para obtener las mejores puntuaciones posibles.

a los “buenos estudiantes” para que la muestra juegue a su favor, reducir el plan curricular, promover la orientación de la enseñanza en función de las pruebas y la reducción de los planes de estudios u orientar la enseñanza en función de las pruebas (*teaching to the test*), prestar solo atención a aquellos alumnos con más probabilidades de aprobar o incluso fomentar las trampas o fraude (UNESCO, 2017d).

Estas prácticas resultan problemáticas por diversos motivos. En primer lugar, constituyen una forma de burlar el sistema que puede desorientar las labores de seguimiento y dar lugar a puntuaciones exageradas que no representen los conocimientos reales de los estudiantes (Koretz, 2017). En segundo lugar, es probable que acaben socavando la calidad y equidad generales de la educación. De hecho, incluso sus efectos sobre los resultados del aprendizaje siguen siendo objeto de debate. Por ejemplo, las conclusiones del *Informe de Seguimiento de la Educación en el Mundo 2017/2018* muestran que los resultados del aprendizaje, evaluados en función de las puntuaciones obtenidas en los PISA entre 2003 y 2015, no han mejorado en los sistemas de rendición de cuentas basados en pruebas. Por último, pueden acarrear la reducción *de facto* o informal del plan de estudios, puesto que los conocimientos, las habilidades, las competencias, las actitudes y los valores que no forman parte de la prueba suelen recibir menos atención⁴⁴.

El caso de la remuneración basada en el mérito

La dinámica consistente en emplear sanciones y recompensas es especialmente visible en el caso de los sistemas de remuneración basados en el desempeño: una política pública que se ha ido consolidando a lo largo del último decenio. En muchos países, las calificaciones de las pruebas de los estudiantes se tienen cada vez más en cuenta en las evaluaciones de los docentes (UNESCO, 2017d) y, en consecuencia, están vinculadas a su remuneración, sus ascensos y su situación laboral (Larsen, 2005). Es el caso, por ejemplo, de los Estados Unidos, Portugal, Chile e Inglaterra, que han probado con diferentes programas de remuneración ligada al desempeño (UNESCO, 2014).

Aunque esto puede motivar al profesorado a esforzarse por mejorar los resultados de sus alumnos, las sanciones impuestas a los docentes sobre la base de las calificaciones de las pruebas o las evaluaciones de los estudiantes pueden tener múltiples consecuencias negativas para la enseñanza, el aprendizaje y la equidad (UNESCO, 2017d). Según el *Informe de Seguimiento de la Educación en el Mundo 2017/2018*, la remuneración basada en el desempeño puede promover un entorno competitivo poco sano, reducir la motivación de los docentes y propiciar la orientación de la enseñanza en función de las pruebas (UNESCO, 2017d). Además, debe tenerse en cuenta que la calidad de la enseñanza no se debe determinar únicamente mediante los resultados obtenidos en las pruebas, sino también mediante los valores y principios de los docentes; su profesionalidad; la medida en que los materiales disponibles; o su capacidad de ofrecer un entorno de aprendizaje positivo y de enseñar de acuerdo con las necesidades individuales de cada alumno. Por tanto, asociar el desempeño de los alumnos

⁴⁴ Las pruebas sobre estas cuestiones siguen siendo objeto de importantes disputas. Hay quien afirma que, conforme a un análisis del contenido de los libros de texto, no existe evidencia de la reducción de los planes de estudios (Ramírez *et al.*, 2018). Otros expertos han señalado que los efectos de la nueva cultura de la evaluación podrían manifestarse de formas más sutiles, como en la distribución del tiempo de instrucción entre las distintas esferas de competencia (Smith, 2019).

con el salario de los docentes puede hacer que los requisitos para ser un “buen profesor” sean muy limitados. Es posible que los docentes frustrados y desmotivados opten por dejar su trabajo, lo que dificultará la contratación y la retención del personal en el cuerpo docente. Se ha documentado esta dinámica en los Estados Unidos, donde la rendición de cuentas basada en pruebas está especialmente afianzada. Como indican Ryan *et al.* (2017), las evaluaciones con grandes repercusiones tienen un efecto negativo sobre la retención del personal docente, debido a los niveles de estrés y agotamiento.

¿Un fenómeno global en aumento?

El uso de las LSLA para la rendición de cuentas está presente en países de todo el mundo y constituye una tendencia generalizada y consolidada. Algunos de los ejemplos más conocidos y afianzados son Australia, donde las LSLA se utilizan para comparar de manera pública las escuelas y recompensar a los docentes (Hardy y Boyle, 2011); los Estados Unidos, donde los resultados de las pruebas se usan para clausurar las escuelas cuyos estudiantes se encuentran por debajo de los criterios de competencias estatales (Darling-Hammond, 2004); Chile, donde la financiación de las escuelas se ve afectada por el desempeño en las evaluaciones de aprendizaje (UNESCO, 2017d); o, más recientemente, la República Unida de Tanzania, donde un proyecto piloto estableció la remuneración basada en el desempeño para los docentes y las escuelas sobre la base de los resultados de las pruebas (Mbiti *et al.*, 2018).

Sin embargo, los datos disponibles sugieren que, en los países en desarrollo, la mayoría de las LSLA se basan en muestreos (IEU, 2019a) y, por consiguiente, no están en condiciones de fundamentar ningún marco de rendición de cuentas con grandes repercusiones. Además, la expansión de los mecanismos de rendición de cuentas no parece avanzar de manera lineal ni irreversible (Verger *et al.*, 2019). Existe evidencia de que, últimamente, algunos países se han esforzado por separar las iniciativas de rendición de cuentas de las LSLA. Es el caso de Dinamarca, que ha adoptado medidas para garantizar que sus alumnos, sus docentes y sus escuelas estén protegidos frente a las medidas punitivas asociadas al desempeño de los alumnos en las evaluaciones (véase **Cuadro 6**). En conjunto, estas tendencias indican que, si bien la rendición de cuentas basada en pruebas constituye un fenómeno en proceso de mundialización, no tiene un carácter *totalizante*.

Estas observaciones tienen importantes repercusiones en el debate en torno a las LSLA. En varios países de ingresos altos, la expansión y la consolidación de pruebas estandarizadas se ha visto acompañada, de manera simultánea o inmediatamente después, por la adopción de instrumentos de rendición de cuentas con grandes repercusiones. En estos casos, normalmente se ha hablado de la pertinencia y los riesgos de las LSLA en tanto que componente de los marcos de rendición de cuentas. Sin embargo, este enfoque no permite estudiar ni debatir el papel ni los efectos que tienen las LSLA en los países de ingresos bajos y medianos. Además, supeditar el debate sobre las evaluaciones al debate sobre la rendición de cuentas con grandes repercusiones conlleva riesgos importantes también cuando se trata

Supeditar el debate sobre las evaluaciones al debate sobre la rendición de cuentas con grandes repercusiones conlleva riesgos importantes también cuando se trata de países de ingresos altos.

de países de ingresos altos. Esto se debe a que algunas de las consecuencias derivadas de la utilización de instrumentos de evaluación (no ligadas a su utilización a efectos de rendición de cuentas) pueden pasar desapercibidas con facilidad.

Cuadro 6. Protegiendo a los alumnos, los docentes y las escuelas frente a las sanciones asociadas al desempeño

Dinamarca lleva a cabo una evaluación a gran escala de base censal centrada en distintas materias y administrada entre los cursos 2º y 8º. La prueba, informatizada y adaptativa (es decir, no hay dos alumnos que hagan la misma prueba) es también una de las distintas herramientas pedagógicas a disposición de los docentes para proporcionarles información diagnóstica y ayudarlos a evaluar el desempeño de los alumnos y a diseñar planes individuales para cada estudiante.

Los resultados son analizados a nivel individual, por grupos de alumnos (agregados por sexo, edad, condición socioeconómica u otros) y por aulas, escuelas, municipios y regiones. Están disponibles de manera pública solo por subgrupos en el plano nacional. Los datos preservan el carácter confidencial y están protegidos por la legislación nacional. Al asegurar la confidencialidad y el anonimato de los alumnos y de otros tipos de informaciones identificativas, la legislación vela por que las escuelas, los docentes y los alumnos estén protegidos ante posibles medidas punitivas asociadas a los resultados de las evaluaciones de los estudiantes.

Fuente: Basado en Shewbridge *et al.*, 2011; OCDE, 2014.

Evaluaciones con pocas repercusiones y enseñanza en función de las pruebas: un debate nuevo

En contraste con la atención prestada a los marcos de rendición de cuentas con grandes repercusiones, la comprensión y reconocimiento de las consecuencias de las evaluaciones con pocas repercusiones son limitados. Por ejemplo, la orientación de la enseñanza en función de las pruebas y la reducción de los planes de estudios se han estudiado por lo general como una consecuencia imprevista de los exámenes con grandes repercusiones, y no de las LSLA destinadas principalmente a ofrecer información sobre el estado general de todo el sistema. Algunos investigadores reconocen que los inconvenientes de la orientación de la enseñanza en función de las pruebas, inherentes a las repercusiones de las LSLA en las políticas, pueden haberse exagerado (Ritzen, 2013). Sin embargo, otros defienden que, incluso cuando las evaluaciones nacionales tienen pocas repercusiones para el alumno, pueden impulsar a los docentes a orientar la enseñanza a las pruebas debido a las consecuencias simbólicas de las LSLA y la percepción de un riesgo para la reputación de la escuela y del docente. Como ya se ha mencionado, los países utilizan cada vez más su posición general en los cuadros de clasificación de evaluaciones internacionales como el proyecto PISA a modo de indicador nacional de progreso y mejora, o de falta de estos⁴⁵. Esto puede suponer un incentivo para

Incluso las evaluaciones con pocas repercusiones podrían tener en última instancia una importancia decisiva a nivel simbólico, fomentando la competencia entre escuelas o llevándolas a esforzarse por definir el grupo que se someterá a las pruebas.

⁴⁵ Véase, por ejemplo, Takayama (2008); Niemann *et al.* (2017); o Baroutsis y Lingard (2017) para obtener información general sobre los casos japonés, alemán y australiano, respectivamente.

que las autoridades de educación remodelen los planes de estudios o adapten las prácticas docentes al contenido o el estilo de las evaluaciones transnacionales con el objeto de mejorar el desempeño de los estudiantes. De forma similar, las evaluaciones nacionales que no están vinculadas de manera formal con incentivos concretos para los docentes y las escuelas en ocasiones se utilizan deliberadamente para guiar la elección escolar y para fomentar la competencia entre escuelas. Existen más probabilidades de que ocurra esto cuando los resultados de las pruebas se desglosan por escuelas y se ponen a disposición del público general, a veces en forma de cuadros de clasificación y *rankings*⁴⁶. De este modo, incluso las evaluaciones con pocas repercusiones podrían tener en última instancia una importancia decisiva a nivel simbólico, fomentando la competencia entre escuelas o llevándolas a esforzarse por definir el grupo que se someterá a las pruebas⁴⁷. Emerge pues un nuevo debate sobre si las evaluaciones transnacionales basadas en muestreos o las evaluaciones nacionales con pocas repercusiones también podrían influir en las estrategias de enseñanza y aprendizaje.

46 De hecho, incluso aunque no haya mecanismos de este tipo vigentes, las escuelas podrían usar sus resultados con fines promocionales.

47 Es más probable que esta dinámica se desarrolle en sistemas educativos que presenten altos niveles de libertad de elección de escuela y que estén organizados en función de una dinámica de mercado. En estos contextos, las escuelas encuentran un mayor incentivo para concentrar sus esfuerzos en mejorar el desempeño escolar en las pruebas estandarizadas, ya que conseguir una reputación en este ámbito seguramente tendrá un efecto determinante sobre el nivel de matrícula (véase, por ejemplo, Falabella, 2014, para el caso chileno).

Reflexiones finales

El potencial de las evaluaciones de aprendizaje a gran escala (LSLA) como factores impulsores del progreso educativo se ha consolidado con firmeza en una bibliografía especializada cada vez más amplia. A lo largo de los últimos decenios, el interés por las LSLA en los planos nacional, regional e internacional no ha parado de crecer hasta un punto de no retorno, en cierto modo. Independientemente de las características que presenten los sistemas educativos en el futuro, las LSLA están aquí para quedarse. No solo está aumentando la cantidad de países que llevan a cabo iniciativas de LSLA, sino que el mercado de las evaluaciones también ha experimentado una transformación significativa. Ha surgido una gran variedad de iniciativas comprometidas con el apoyo a la realización de evaluaciones, con ayuda de los organismos internacionales y regionales de desarrollo. Por consiguiente, el panorama de las evaluaciones ha adquirido un carácter más complejo y diverso, y, con esta transformación, han surgido nuevas dificultades.

Como se explicó anteriormente, las LSLA poseen un gran potencial, pero las ventajas deben sopesarse en comparación con los posibles riesgos. Algunos de estos riesgos se derivan directamente de su diseño, mientras que otros son consecuencia de los usos inapropiados de los resultados. Si bien dichas limitaciones no hacen que las evaluaciones sean inservibles ni intrínsecamente problemáticas, existen motivos para aumentar la vigilancia con el fin de evitar posibles usos inadecuados. Es necesario diseñar las herramientas de un modo más cuidadoso y equilibrado, y utilizar los resultados generados de una forma más moderada para fundamentar correctamente las políticas públicas.

Además, parece que, cada una frecuencia creciente, se dan por sentadas las LSLA en la mayoría de los sistemas educativos. Sin embargo, está aumentando la consciencia de que la repercusión final de las LSLA depende fundamentalmente de *cómo* las utilizemos, como mínimo en igual medida que el diseño y los métodos de recopilación de datos empleados. Aun así, las investigaciones y los debates sobre distintas cuestiones relativas al uso de datos siguen siendo insuficientes. En el presente apartado, presentamos algunas de las esferas de indagación que se beneficiarían de una mayor atención, reflexión e investigación.

El cumplimiento de las promesas de las evaluaciones

Evaluaciones en contexto. Según está previsto, las LSLA deben atender a una gran variedad de propósitos y usos. Sin embargo, existe una comprensión limitada sobre *cómo* y *bajo qué condiciones* pueden cumplir de manera eficaz con estas promesas, es decir, sobre cómo pueden fundamentar con eficacia los cambios políticos o las prácticas para, en última instancia, mejorar la calidad y la equidad de la educación. Es necesario, por tanto, mejorar nuestra comprensión sobre el efecto que tienen sobre las LSLA los factores ligados al contexto económico, institucional y político.

Entender dicho efecto es de vital importancia en vista de la reciente evidencia sobre los posibles efectos secundarios y las consecuencias imprevistas de las evaluaciones. Existen cada vez más investigaciones que documentan cómo las LSLA pueden fomentar comportamientos poco deseables o incluso oportunistas entre diversas partes interesadas. Por consiguiente, es necesario seguir investigando para entender qué evaluaciones y condiciones podrían aumentar la calidad y la equidad de la educación, y cuáles podrían resultar contraproducentes.

Más allá de las evaluaciones de aprendizaje a gran escala. Como se ha señalado a lo largo del documento, es posible que sea necesario ajustar en cierta medida las LSLA para que se adecúen a un programa equitativo y de calidad. Sin embargo, las LSLA no puedan cumplir *todos* los propósitos y objetivos que se esperan de un sistema de evaluación. Es necesario seguir reflexionando sobre el papel que pueden desempeñar otras modalidades de evaluación, como las evaluaciones en el aula con fines formativos y diagnósticos, las evaluaciones de portafolio y las evaluaciones basadas en observación o por pares. De hecho, estas evaluaciones pueden encajar mejor a la hora de medir esferas de aprendizaje relevantes que no se prestan a comparaciones a nivel de sistema o entre países. Por lo tanto, es necesario prestar mayor atención a otras formas de evaluación para identificar las maneras más apropiadas de reflejar todas las competencias que se consideren valiosas.

De forma similar, si bien las evaluaciones gubernamentales o intergubernamentales han recibido una atención considerable en los últimos decenios, existen otras modalidades emergentes de LSLA que no son tan conocidas. Destacan, en particular, las evaluaciones impulsadas por los ciudadanos y las encuestas de hogares que contienen un componente de evaluación. El interés por estos temas sigue limitándose, en gran medida, a las comunidades de práctica especializadas que están a cargo de ellos. De esta manera, las oportunidades y riesgos que conllevan estas evaluaciones rara vez son objeto de estudios generales. Sin embargo, estudiar estas formas alternativas de evaluación puede revelar maneras de subsanar o compensar algunas de las limitaciones de las LSLA convencionales. En concreto, tienen el potencial de garantizar un mayor grado de equidad e inclusión al incorporar idiomas adicionales y llegar a poblaciones remotas o a niños sin escolarizar. Por lo tanto, podría ser de utilidad reflexionar de manera fundamentada sobre su potencial y sus limitaciones para promover el diseño de marcos de evaluación eficaces.

El futuro de las evaluaciones

Es necesario poner remedio a las posibles limitaciones planteadas en el apartado anterior para poder emplear las LSLA de manera óptima al servicio de la calidad y la equidad de la educación. No obstante, la mayoría de las dificultades asociadas a las LSLA se manifestarán en el futuro. Las posibles cuestiones sobre las que debemos seguir reflexionando son las siguientes:

El papel del sector privado. Existen cada vez más organizaciones que forman parte del campo de las evaluaciones, algunas de las cuales resultan todavía poco conocidas o son raramente examinadas en profundidad. Hay una gran variedad de actores invisibles que desempeñan una función esencial en el proceso de evaluación. Cabe destacar el caso de la mayoría de los contratistas, quienes asumen una parte considerable de las tareas de evaluación y, sin embargo, están sujetos a niveles relativamente bajos de visibilidad y escrutinio público. Dado el estatus

privado de muchos de ellos, es frecuente que estos actores intermediarios sigan siendo puntos ciegos. Por consiguiente, para describir el nuevo panorama de las evaluaciones es necesario prestar especial atención al papel del sector privado.

Un cambio en el entorno de aprendizaje. En la actualidad, la comprensión de la evolución de la naturaleza del aprendizaje en los últimos años, especialmente el modo en que el aprendizaje se ha visto afectado por el cambio tecnológico, es limitada. Los cuestionarios de antecedentes que suelen acompañar las evaluaciones han prestado una atención reducida e irregular a las prácticas de aprendizaje en la infancia y la juventud, así como a las actitudes con respecto al aprendizaje dentro y fuera de las escuelas. Se necesitan más iniciativas que incorporen estas dimensiones en los cuestionarios de antecedentes, de tal forma que las evaluaciones ayuden a entender mejor los nuevos entornos de aprendizaje en constante cambio en los que están inmersos los alumnos del siglo XXI.

La ética de las evaluaciones. La disponibilidad de un conjunto cada vez más amplio de datos de las evaluaciones plantea una serie de cuestiones éticas relacionadas con el consentimiento y la transparencia, la propiedad de los datos, y la seguridad y el anonimato. Si bien hay un creciente interés por las normas éticas necesarias para la reunión y el uso de los datos, este todavía se encuentra en una etapa incipiente. Es necesario seguir investigando y reflexionando para comprender mejor dónde residen los posibles riesgos y dificultades, así como para aclarar los diferentes niveles de responsabilidad. También es necesario definir un entendimiento común de las buenas prácticas de gestión de datos que pueda subsanar de forma adecuada las nuevas dificultades planteadas por la aparición de *big data* y la inteligencia artificial.

Datos de las evaluaciones como big data. En la actualidad, el término *big data* es uno de los más sobreutilizados y, al mismo tiempo, menos comprendidos. Aunque tienen un papel destacado en los debates públicos y especializados, todavía no está claro si los datos de las evaluaciones se pueden clasificar dentro de esta categoría. El aumento de las evaluaciones informatizadas y las pruebas adaptativas, y la mayor dependencia de algoritmos para analizar y extraer conclusiones a partir de los datos de evaluación no se han visto acompañados de una reflexión o un debate sobre las repercusiones que dichas tendencias tienen en las personas evaluadas ni en los sistemas educativos en general, e incluso en nuestros conceptos de calidad y equidad de la educación. Por consiguiente, mejorar el conocimiento de la relación entre la *big data* y los datos de evaluación es un paso clave para asegurar que los datos siguen estando al servicio de las partes interesadas en la educación, y no al contrario.

En vista de la rápida transformación que experimentan la educación y el aprendizaje, entender de una manera más amplia y precisa las nuevas dificultades se ha convertido en una necesidad imperiosa. Debemos responder al llamamiento a una reflexión crítica y un diálogo en profundidad si queremos cumplir nuestro compromiso colectivo de garantizar un aprendizaje eficaz y pertinente para todos.

Referencias

- ACER-IEU. 2017. *The Principles of Good Practice in Learning Assessment*. Melbourne: ACER.
- Addey, C. 2014. *Why do countries join international literacy assessments? An Actor-Network Theory analysis with case studies from Lao PDR and Mongolia*. Facultad de Educación y Aprendizaje Permanente del Reino Unido: Universidad de Anglia Oriental.
- Addey, C. 2018. "The assessment culture of international organizations: from philosophical doubt to statistical certainty through the appearance and growth of international large-scale assessments" en C. Alarcón y M. Lawn (eds.), *Assessment Cultures: Historical perspectives*. Fráncfort del Meno: Peter Lang.
- Addey, C. y Sellar, S. 2019. ¿Vale la pena? Motivaciones para participar (o no) en evaluaciones internacionales del aprendizaje. *Education Investigación y Prospectiva en Educación*, No. 24. París, UNESCO.
- Addey, C., Sellar, S., Steiner-Khamsi, G., Lingard, B. y Verger, T. 2017. "The rise of international large-scale assessments and rationales for participation". *Compare: A Journal of Comparative and International Education*.
- Agencia Sueca de Cooperación Internacional para el Desarrollo (ASDI). 2018. Página "Education" del sitio web de la ASDI. <https://www.sida.se/English/how-we-work/our-fields-of-work/education/> (consultado en julio de 2018).
- Alianza Mundial para la Educación. 2016. *Plan Estratégico para 2020 de la Alianza Mundial para la Educación*. Washington D. C.: Alianza Mundial para la Educación.
- Alianza Mundial para la Educación. 2018. *Plan Estratégico para 2020 de la Alianza Mundial para la Educación*. Washington D. C.: Alianza Mundial para la Educación.
- Anderson, J. A. 2005. *Accountability in education*. París: Instituto Internacional de Planeamiento de la Educación de la UNESCO.
- Baird, J. A., Johnson, S., Hopfenbeck, T. N., Isaacs, T., Sprague, T., Stobart, G. y Yu, G. 2016. "On the supranational spell of PISA in policy". *Educational Research*, vol. 58, núm. 2, págs. 121 a 138.
- Banco Mundial. 2011. *Aprendizaje para todos: invertir en los conocimientos y las capacidades de las personas para fomentar el desarrollo*. Washington D. C.: Banco Mundial.
- Banco Mundial. 2018. *Learning to Realize Education's Promise. The World Development Report 2018*. Washington D. C.: Banco Mundial.
- Barnes, S., Lyonette, C., Atfield, G. y Owen, D. 2016. *Teachers' Pay and Equality: A Literature Review – Longitudinal Research into the Impact of Changes to Teachers' Pay on Equality in Schools in England*. Warwickshire, Reino Unido: Instituto para la Investigación del Empleo de Warwick.
- Baroutsis, A. y Lingard, B. 2017. "Counting and comparing school performance: an analysis of media coverage of PISA in Australia, 2000–2014". *Journal of Education Policy*, vol. 32, núm. 4, págs. 432 a 449.
- Barrett, A. M. 2016. "Measuring Learning Outcomes and Education for Sustainable Development: the new education development goal" en W. C. Smith, y W. C. Smith (ed.), *The Global Testing Culture; shaping education policy, perceptions, and practice*. Reino Unido: Symposium Books.

- Beggs, C. 2016. *Overview of EGRA and EGMA*. Alianza Mundial para el Seguimiento del Aprendizaje. http://gaml.uis.unesco.org/wp-content/uploads/sites/2/2018/08/4_USAID_20160511.pdf (consultado en febrero de 2019).
- Benavot, A. y Köseleci, N. 2015. *Seeking Quality in Education: The Growth of National Learning Assessments, 1990-2013*. Documento encargado para el *Informe de Seguimiento de la Educación para Todos en el Mundo 2015*.
- Benson, C. y Wong, K. 2015. "Development discourse on language of instruction and literacy: Sound policy and Ubuntu, or lip service?" *Reconsidering Development*, vol. 4, núm 1, págs. 1 a 16.
- Best, M., Knight, P., Lietz, P., Lockwood, C., Nugroho, D. y Tobin, M. 2013. *The impact of national and international assessment programmes on education policy, particularly policies regarding resource allocation and teaching and learning practices in developing countries*. Londres: Centro de Información y Coordinación de Pruebas para Políticas y Prácticas (EPPI-Centre) de la Unidad de Investigaciones en Ciencias Sociales, Instituto de Educación, Universidad de Londres.
- Biesta, G. 2009. "Good education in an age of measurement: On the need to reconnect with the question of purpose in education". *Educational Assessment, Evaluation and Accountability*, vol. 21, núm. 1, págs. 33 a 46.
- Breakspear, S. 2014. *How does PISA shape education policy making? Why how we measure learning determines what counts in education*. Australia: Center for Strategic Education.
- Brese, F., Jung, M., Mirazchiyski, P., Schulz, W. y Zuehlke, O. 2011. *ICCS 2009 User Guide for the International Database – Supplement 5*. Ámsterdam: Secretaría de la IEA.
- Brink, S. 2018. *Learning from promising practices in ensuring effective use of large-scale learning assessment data for education policy and practice: Developing evidence from national and international LSLA for policy and practice*. Presentación realizada en el taller TALENT celebrado en noviembre de 2018, Dakar (Senegal). http://www.education2030-africa.org/images/Developing_evidence_from_national_and_international_LSLA_for_policy_and_practice.pdf (consultado en febrero de 2019).
- Clarke, M. 2012. *What Matters Most for Student Assessment Systems: A Framework Paper*. Documento de trabajo núm. 1 del Planteamiento sistémico para mejorar los resultados educativos (SABER). Washington D. C.: Grupo Banco Mundial.
- Clarke, M. 2016. *Value for money: What do developing countries really get from participating in international assessment programs?* Presentación realizada el 13 de enero de 2016, Washington D. C. http://siteresources.worldbank.org/1818SOCIETY/Resources/Value_for_Money_Clarke_13Jan2016.pdf (consultado en mayo de 2019).
- Clarke, P. 2017. *Making use of assessments for creating stronger education systems and improving teaching and learning*. Documento de antecedentes para el *Informe de Seguimiento de la Educación en el Mundo 2017/2018*.
- Cresswell, J., Schwantner, U. y Waters, C. 2015. *A Review of International Large-Scale Assessments in Education: Assessing Component Skills and Collecting Contextual Data*, PISA, Banco Mundial, Washington D. C./OECD Publishing, París: OCDE.
- Colclough, C. 2012. "Education, poverty and development—mapping their interconnections". *Comparative Education*, vol. 48, núm. 2, págs. 135 a 148.
- Comisión de Educación (Comisión Internacional sobre la Financiación de las Oportunidades de Educación Mundial). 2016. "The Learning Generation: investing in education for a changing world". *Comparative Education Review*, vol. 61, núm. 2, págs. 214 a 217.

- Comisión Especial sobre Métricas de los Aprendizajes. 2013. *Hacia un aprendizaje universal: recomendaciones de la Comisión Especial sobre Métricas de los Aprendizajes*. Montreal y Washington, D. C.: Instituto de Estadística de la UNESCO y Centro de Educación Universal de Brookings Institution.
- Darling-Hammond, L. 2004. "Standards, accountability and school reform". *Teachers College Record*, vol. 106, núm. 6, págs. 1047 a 1085.
- De Coster, I. y Sigalas, E. 2017. *Informe de Eurydice: La educación para la ciudadanía en el contexto escolar europeo 2017*. Comisión Europea: Agencia Ejecutiva en el ámbito Educativo, Audiovisual y Cultural.
- Departamento de Relaciones Exteriores y Comercio de Australia 2015. *Strategy for Australia's aid investments in education 2015–2020*. Australia: Departamento de Relaciones Exteriores y Comercio.
- Departamento de Relaciones Exteriores y Comercio del Canadá. 2014. *Canada's International Education Strategy: Harnessing out knowledge advantage to drive innovation and prosperity*. Ottawa: Departamento de Relaciones Exteriores y Comercio del Canadá.
- DFID. 2018. *DFID Education Policy: Get Children Learning*. Londres: DFID.
- Díaz, M. y Flores, G. 2008. *SERCE, Resultados Nacionales*. Ciudad de México: Instituto Nacional para la Evaluación de la Educación.
- Dixon, R., Arndt, C., Mullers, M., Vakkuri, J., Engblom-Pelkkala, K. y Hood, C. 2013. "A Lever for Improvement or a Magnet for Blame? Press and Political Responses to International Educational Rankings in Four EU Countries". *Public Administration*, vol. 91, núm. 2, págs. 484 a 505.
- Dolata, S. 2011. *Are there any disparities between girls and boys in the response of the education sector of hiv and aids? Assessment of educational HIV/AIDS prevention programmes applied by SACMEQ III countries*. Documento presentado en el Foro de Políticas Educativas relativas a la Igualdad de Género del IIPE, 3 y 4 de octubre de 2011, París (Francia).
- Earle, D. 2010. *How can tertiary education deliver better value to the economy?* Documento ocasional sobre la enseñanza terciaria, agosto de 2010. Nueva Zelanda: Gobierno de Nueva Zelanda.
- Engel, L. C. y Rutkowski, D. 2018. "Pay to play: what does PISA participation cost in the US?" *Discourse: Studies in the Cultural Politics of Education*. Routledge.
- Falabella, A. 2014. "The Performing School: The Effects of Market & Accountability Policies". *Education Policy Analysis Archives*, vol. 22, núm. 70.
- Fenwick, T. y Edwards, R. 2014. "Networks of knowledge, matters of learning, and criticality in higher education". *Higher Education*, vol. 67, núm. 1, págs. 35 a 50.
- Ferrer, G. y Fiszbein, A. 2015. *What has happened with learning assessment systems in Latin America? Lessons from the Last Decade of Experience*. The Dialogue: Leadership for the Americas. Comisión para la Educación de Calidad para Todos: documento de antecedentes. Washington D. C.: Grupo Banco Mundial.
- Feuer, M. 2012. *No country left behind: rhetoric and reality of international large-scale assessment*. William Angoff Memorial Lecture Series. Nueva Jersey: Educational Testing Service.
- Fischman, G. y Topper, A. M. 2017. *An Examination of the Influence of International Large Scale Assessments and Global Learning Metrics on National School Reform Policies*. Facultad de Magisterio Mary Lou Foulton. Universidad Estatal de Arizona.
- Fiske, E.B. 2000. *Evaluación del aprovechamiento escolar. Educación para Todos: Situación y tendencias 2000*. París: UNESCO, París. Secretaría del Foro sobre Educación para Todos

- Friedman, T., Schwantner, U., Spink, J., Tabata, N. y Waters, C. 2016. *Improving Quality Education and Children's Learning Outcomes and Effective Practices in the Eastern and Southern Africa Region: Main Report*. Informe para la Oficina Regional para el África Oriental y Meridional (ESARO) del UNICEF. Camberwell, Victoria: ACER Press.
- Fullan, M. 2010. "Positive pressure". A. Hargreaves, A. Lieberman, M. Fullan y D. Hopkins (eds). En *Second international handbook of educational change* (págs. 119 a 130). Springer, Dordrecht.
- Goldstein, H. y Thomas, S. 2008. "Reflections on the international comparative surveys debate". *Assessment in Education: Principles, Policy & Practice*, vol. 15, núm. 3, págs. 215 a 222.
- Gorard, S. 2001. "International Comparisons of School Effectiveness: the second component of the 'crisis account' in England?" *Comparative Education*, vol. 37, núm. 3, págs. 279 a 296.
- Gorur, R. y Wu, M. 2014. "Leaning too far? PISA, policy and Australia's 'top five' ambitions". *Discourse: Studies in the Cultural Politics of Education*, vol. 36, núm. 5, págs. 647 a 664.
- Greaney, V. 2008. *Evaluación de los niveles nacionales de rendimiento académico*. Washington D. C.: Banco Mundial.
- Hamilton, L.S., Stecher, B.M. y Klein, S.P. 2002. *Making Sense of TBA in Education*. Santa Mónica, California: Rand Corporation.
- Hamilton, M. 2017. "How International Large-Scale Skills Assessments engage with national actors: mobilising networks through policy, media and public knowledge". *Critical Studies in Education*, vol. 58, núm. 3, págs. 280 a 294.
- Hannon, V., Patton, A. y Temperley, J. 2011. *Developing an Innovation - Ecosystem for Education*. Indianápolis, CISCO.
- Hanushek, E. A. y Woessmann, L. 2015. *The Knowledge Capital of Nations: Education and the Economics of Growth*. Cambridge, Massachusetts: MIT Press.
- Hardy, I. y Boyle, C. 2011. "My school? Critiquing the abstraction and quantification of education". *Asia Pacific Journal of Teacher Education*, vol. 39, núm. 3, págs. 211 a 222.
- IEA. 2015. "TIMSS 2015 Context Questionnaires". <https://timssandpirls.bc.edu/timss2015/questionnaires/index.html> (consultado en mayo de 2019).
- IEA. 2016. "PIRLS 2016 Context Questionnaires". <https://timssandpirls.bc.edu/pirls2016/questionnaires/index.html> (consultado en mayo de 2019).
- IEA. 2019. "National Study Reports". http://pub.iea.nl/national_reports.html (consultado en mayo de 2019).
- Iniciativa Mundial de Educación del Secretario General de las Naciones Unidas. 2012-2016. *Iniciativa Mundial La Educación ante Todo*.
- Instituto de Estadística de la UNESCO. 2016a. *The cost of not assessing learning outcomes*. Documento informativo núm. 26. Montreal: Instituto de Estadística de la UNESCO.
- Instituto de Estadística de la UNESCO. 2016b. *Experiencias exitosas en la evaluación oral de la lectura*. Montreal: Instituto de Estadística de la UNESCO.
- Instituto de Estadística de la UNESCO. 2018a. *The impact of large-scale learning assessments*. Montreal: Instituto de Estadística de la UNESCO.
- Instituto de Estadística de la UNESCO. 2018b. *Costs and Benefits of Different Approaches to Measuring the Learning Proficiency of Students (SDG Indicator 4.1.1)*. Documento informativo núm. 53. Montreal: Instituto de Estadística de la UNESCO.

- Instituto de Estadística de la UNESCO. 2018c. *SDG4 Data Digest. Data to Nurture Learning*. Montreal: Instituto de Estadística de la UNESCO.
- Instituto de Estadística de la UNESCO. 2019a. "Database". <http://uis.unesco.org/> (consultado en enero de 2019).
- Instituto de Estadística de la UNESCO. 2019b. "Glossary". <http://uis.unesco.org/en/glossary> (consultado en enero de 2019).
- Instituto Internacional de Planeamiento de la Educación de la UNESCO (IIPE) 2017. *Planificar la educación, construir el futuro. 10a estrategia de medio término 2018-2021*. París: IIPE de la UNESCO.
- Kamens, D. y McNeely, C. 2010. "Globalization and the Growth of International Educational Testing and National Assessment". *Comparative Education Review*, vol. 54, núm. 1, págs. 5 a 25.
- Kellaghan, T. y Greaney, V. 2001. *Using assessment to improve the quality of education*. París: IIPE.
- Kellaghan, T., Bethell, G. y Ross, J. 2011. *National and international assessments of student achievement. Guidance Note: A DFID Practice Paper*. Londres: Departamento para el Desarrollo Internacional del Reino Unido (DFID).
- King, K. 2017. "Lost in translation? The challenge of translating the global education goal and targets into global indicators". *Compare: A journal of comparative and international education*, vol. 47, núm. 6, págs. 801 a 817.
- Komatsu, H. y Rappleye, J. 2017. "A new global policy regime founded on invalid statistics? Hanushek, Woessmann, PISA, and economic growth". *Comparative Education*, vol. 53, núm. 2, págs. 166 a 191.
- Koretz, D. 2017. *The testing charade. Pretending to make schools better*. Chicago y Londres: The University of Chicago Press.
- Larsen, M. A. 2005. "A Critical Analysis of Teacher Evaluation Policy Trends". *Australian Journal of Education*, vol. 49, núm. 3, págs. 292 a 305.
- Lingard, B. y Grek, S. 2007. *The OECD, indicators and PISA: an exploration of events and theoretical perspectives. ESRC/ESF Research Project on Fabricating Quality in Education Working Paper 2*. Universidad de Edimburgo.
- Lingard, B. y Sellar, S. 2013. "'Catalyst data': Perverse systemic effects of audit and accountability in Australian schooling". *Journal of Education Policy*, vol. 28, núm. 5, págs. 634 a 656.
- Lockheed, M. 2015. *Why do countries participate in international large-scale assessments? The Case of PISA*. Policy Research Working Paper. Grupo Banco Mundial.
- Mbiti, I., Romero, M. y Schipper, Y. 2018. *The Challenge of Designing Effective Teacher Performance Pay Programs: Experimental Evidence from Tanzania*. Documento de trabajo. http://barrett.dyson.cornell.edu/NEUDC/paper_613.pdf
- Medición del Aprendizaje en Primaria para el Sudeste Asiático (SEA-PLM), Organización de Ministros de Educación de Asia Suroriental (OMEASO), UNICEF y ACER. 2017. *Assessment Framework. Revised draft document incorporating feedback from the Domain Technical Review Teams*.
- Ministerio de Educación, Finlandia. 2006. *Education Strategy for Finland's Development Cooperation*. Finlandia: Ministerio de Educación.
- Mitra, D., Mann, B. y Hlavacik, M. 2016. "Opting Out: Parents Creating Contested Spaces to Challenge Standardized Tests". *Education Policy Analysis Archives*, vol. 24, núm. 31.
- Montoya, S. 2016. *The Cost of Ignorance Revisited: A Reply*. <https://norrag.wordpress.com/2016/02/08/measuring-learning-the-cost-of-ignorance/> (consultado en junio de 2017).

- Niemann, D., Martens, K. y Teltemann, J. 2017. "PISA and its consequences: Shaping education policies through international comparisons". *European Journal of Education*, vol. 52, núm. 2, págs. 175 a 183.
- Nikel, J. y Lowe, L. 2010. "Talking of fabric: a multi-dimensional model of quality in education". *Compare: A journal of comparative and international education*, vol. 40, núm. 5, págs. 589 a 605.
- Organismo Noruego de Cooperación para el Desarrollo (NORAD). 2018. Página web "Education". <https://norad.no/en/front/thematic-areas/education/> (consultado en julio de 2018).
- OCDE. 2009. *PISA Technical Report*. París: OCDE.
- OCDE. 2010. *The high cost of low educational performance: The long-run economic impact of improving PISA outcomes*. París: OCDE.
- OCDE. 2011. *Improving access to PISA for students with disabilities and other special education needs*. París: OCDE.
- OCDE. 2013. *Synergies for Better Learning. An international perspective on evaluation and assessment. OECD Reviews of Evaluation and Assessment in Education*. París: OCDE.
- OCDE. 2014. *Education Policy Outlook: Denmark*. París, OCDE.
- OCDE. 2015. "Correr en días de calor. Introducción". Elemento de la prueba del PISA 2015. <http://www.oecd.org/pisa/PISA2015Questions/platform/index.html?user=&domain=SCI&unit=S623-RunningInHotWeather&lang=esp-esp> (versión original consultada en diciembre de 2018).
- OCDE. 2016a. *PISA. Resultados en foco*. París: OCDE.
- OCDE. 2016b. Folleto "PISA para el Desarrollo". https://www.oecd.org/pisa/pisa-for-development/PISA-D_brochure_2016_ESP.pdf (versión original consultada en agosto de 2017).
- OCDE. 2017. *PISA 2015 Collaborative Problem-Solving Framework*. París: OCDE.
- OCDE. 2018a. *The Future of Education and Skills: Education 2030*. París: OCDE.
- OCDE. 2018b. "Data Management Procedures". En *PISA for Development Technical Report*. París: OCDE. <http://www.oecd.org/pisa/pisa-for-development/pisafordevelopment2018technicalreport/PISA-D%20TR%20Chapter%2010%20-%20DataManagement%20-%20final.pdf> (consultado en mayo de 2019).
- Oficina Internacional de Educación de la UNESCO (OIE). 2013. *El Aprendizaje en la agenda para la educación y el desarrollo después de 2015*.
- Oficina de la UNESCO en Santiago de Chile (Oficina Regional de Educación para América Latina y el Caribe (OREALC)/UNESCO Santiago). 2007. *Educación de calidad para todos: un asunto de derechos humanos*. Documento de discusión sobre políticas educativas en el marco de la II Reunión Intergubernamental del Proyecto Regional de Educación para América Latina y el Caribe (EPT/ PRELAC). Documento de antecedentes. Santiago de Chile: Oficina Regional de Educación para América Latina y el Caribe de la UNESCO.
- Parker, R. y Fraillon, J. 2016. *Southeast Asia Primary Learning Metrics (SEA-PLM) - Global Citizenship Domain Assessment Framework*. Melbourne: ACER.
- PASEC. 2015. *PASEC 2014 Education System Performance in Francophone Sub-Saharan Africa: Competencies and learning factors in primary education*. Dakar: Programa de Análisis de los Sistemas Educativos de la CONFEMEN.
- Pizmony-Levy, O. 2013. *Back to the Future in International Assessments*. http://blogs.edweek.org/edweek/assessing_the_assessments/2014/04/back_to_the_future_how_international_large-scale_assessments_came_about.html (consultado en agosto de 2017).

- Pizmony-Levy, O., Harvey, J., Schmidt, W.H., Noonan, R., Engel, L., Feuer, M.J., Braun, H., Santorno, Rotberg, I.C., Ash, P., Chatterji, M. y Torney-Purta, J. 2014. "On the merits of, and myths about, international assessments". *Quality Assurance in Education*, vol. 22, núm. 4, págs. 319 a 338.
- Pizmony-Levy, O. y Green Saraisky, N. 2016. *Who opts out and why? Results from a national survey on opting out of standardized tests*. Nueva York: Facultad de Magisterio de la Universidad de Columbia.
- Ramirez, F. O., Schofer, E. y Meyer, J.W. 2018. "International Tests, National Assessments, and Educational Development (1970–2012)". *Comparative Education Review*, vol. 62, núm. 3, págs. 344 a 364.
- Raudonyte, I. 2019. *Use of learning assessment data in education policy-making*. Documento de trabajo del IIEP París: IIEP.
- Red de Acción Ciudadana por el Aprendizaje (PAL). 2019. "Nuestro crecimiento". <https://palnetwork.org/our-growth/?lang=es> (versión original consultada en febrero de 2019).
- Ritzen J. 2013. "International large-scale assessments as change agents". Von Davier M, Gonzalez E, Kirsch I et al. (eds). *The Role of International Large-Scale Assessments: Perspectives from Technology, Economy, and Educational Research*. Dordrecht (Países Bajos): Springer, págs. 13 a 24.
- Rowe, E. 2017. "Australia Country Case Study". Documento de antecedentes sobre la práctica del muestreo selectivo en Australia escrito para el *Informe de Seguimiento de la Educación en el Mundo 2017/2018*.
- Rutkowski, L. y Rutkowski, D. 2016. "A call for a more measured approach to reporting and interpreting PISA results". *Educational Researcher*, vol. 45, núm. 4, págs. 252 a 257.
- Ryan, S.V., von der Embse, N.P., Pendergast, L.L., Saeki, E., Segool, N. y Schwing, S. 2017. "Leaving the teaching profession: The role of teacher stress and educational accountability policies on turnover intent". *Teaching and Teacher Education*, vol. 66, págs. 1 a 11.
- Sahlberg, P. 2007. "Education policies for raising student learning: The Finnish approach". *Journal of Education Policy*, vol. 22, núm. 2, págs. 147 a 171.
- Sayed, Y., Ahmed, R. y Mogliacci, R. 2018. "The 2030 Global Education Agenda and the SDGs: Process, Policy and Prospects". En A. Verger, M. Novelli, & H. K. Altinyelken (eds.). *Global Education Policy and International Development (2nd edition)*. Londres/Nueva York: Bloomsbury, págs. 185 a 208.
- Shewbridge, C., Jang, E., Matthews, P. y Santiago, P. 2011. *OECD Reviews of Evaluation and Assessment in Education*. París: OCDE.
- Siddiq, F., Hatlevik, O. E., Olsen, R. V., Throndsen, I. y Scherer, R. 2016. "Taking a future perspective by learning from the past – A systematic review of assessment instruments that aim to measure primary and secondary school students' ICT literacy". *Educational Research Review*, vol. 19, págs. 58 a 84.
- Singal y Sabates. 2016. "Access and learning are equally important for children with disabilities". Sitio web de la Alianza Mundial para la Educación. <https://www.globalpartnership.org/blog/access-and-learning-are-equally-important-children-disabilities> (consultado en mayo de 2019).
- Smith, W.C. 2016. "An Introduction to the Global Testing Culture". W. C. Smith (ed.). En *The Global Testing Culture; shaping educational policy, perceptions and practice*. *Oxford Studies in Comparative Education Series: Symposium Books*.
- Smith, W.C. 2019. "Evaluating the Global Testing Culture: A Response to Ramirez and Colleagues". *Comparative Education Review*, vol. 63, núm. 1, págs. 121 a 126.
- Steiner-Khamsi, G. 2003. "The politics of league tables". *Journal of Social Science Education*. OJSSE 1-2003: Civic Education.

- Steiner-Khamsi, G., Appleton, M. y Vellani, S. 2017. "Understanding business interests in international large-scale student assessments: a media analysis of *The Economist*, *Financial Times*, and *Wall Street Journal*". *Oxford Review of Education*, vol. 44, núm. 2, págs. 190 a 203.
- Takayama, K. 2008. "The politics of international league tables: PISA in Japan's achievement crisis debate". *Comparative Education*, vol. 44, núm. 4, págs. 387 a 407.
- Tawil, S., Akkari, A. y Macedo, B. 2012. Más allá del laberinto conceptual: la noción de la calidad en la educación. *Education Investigación y Prospectiva en Educación*, No. 02. París, UNESCO.
- Tikly, I. y Barrett, A.M. 2007. *Education Quality; Research priorities and approaches in the global era*. Bristol: Universidad de Bristol.
- Tobin, M., Lietz, P., Nugroho, D., Vivekanandan, R. y Nyamkhuu, T. 2015. *Using large-scale assessments of students' learning to inform education policy Insights from the Asia-Pacific region*. ACER y Oficina de la UNESCO en Bangkok: ACER.
- Torney-Purta, J. y Amadeo, J.A. 2013. "International large-scale assessments: Challenges in reporting and potentials for secondary analysis". *Research in Comparative and International Education*, vol. 8, núm. 3, págs. 248 a 258.
- UNESCO, UNICEF, Brookings Institution y Banco Mundial. 2017. *MELQO. Sinopsis. Medición de la Calidad y los Resultados del Aprendizaje Temprano*. París: UNESCO.
- UNESCO. 1974. *Recomendación sobre la educación para la comprensión, la cooperación y la paz internacionales y la educación relativa a los derechos humanos y las libertades fundamentales*. En *Actas de la Conferencia General. 18ª reunión. París, 17 de octubre a 23 de noviembre de 1974*, vol. 1. Resoluciones París: UNESCO.
- UNESCO. 1996. *La educación encierra un tesoro. Informe a la UNESCO de la Comisión Internacional sobre Educación para el Siglo XXI (Compendio)*. Presidente: J. Delors. París: UNESCO.
- UNESCO. 2014. *Informe de Seguimiento de la EPT en el Mundo 2013/2014. Enseñanza y aprendizaje: lograr la calidad para todos*. París: UNESCO.
- UNESCO. 2015a. *Replantear la educación: ¿Hacia un bien común mundial?* París: UNESCO.
- UNESCO. 2015b. *Informe de Seguimiento de la EPT en el Mundo 2015. La Educación para Todos, 2000-2015: logros y desafíos*. París: UNESCO.
- UNESCO. 2016a. *Desglosar el Objetivo de Desarrollo Sostenible 4: Educación 2030*. París: UNESCO.
- UNESCO. 2016b. *Informe de Seguimiento de la Educación en el Mundo 2016: La educación al servicio de los pueblos y el planeta*. París: UNESCO.
- UNESCO. 2016c. *Declaración de Incheon y Marco de Acción: Hacia una Educación Inclusiva y Equitativa de Calidad y un Aprendizaje a lo Largo de la Vida para Todos*. París: UNESCO.
- UNESCO. 2017a. *Analyzing and utilizing assessment data for better learning outcomes*. París y Bangkok: UNESCO.
- UNESCO. 2017b. *Guía para asegurar la inclusión y la equidad en la educación*. París: UNESCO.
- UNESCO. 2017c. *Evaluación del aprendizaje en la UNESCO: Garantía de un aprendizaje efectivo y relevante para todas las personas*. París: UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000260325_spa
- UNESCO. 2017d. *Rendir cuentas en el ámbito de la educación: cumplir nuestros compromisos. Informe de Seguimiento de la Educación en el Mundo*. París: UNESCO.
- UNESCO. 2017e. *Large-scale Assessment Data and Learning Outcomes: Linking Assessments to Evidence-based Policy Making and Improved Learning*. Bangkok: UNESCO.

- UNESCO. 2017f. *SDG 4 Framework for Action: Symposium on Policy implications of learning assessments to enhance educational quality in the Gulf Cooperation Council States*. Doha: UNESCO.
- UNESCO. 2018. *Building tomorrow's digital skills - what conclusions can we draw from international comparative indicators?* Documentos de trabajo de la política educativa núm. 6. París: UNESCO.
- Ungerleider, C. 2006. "Reflections on the use of large-scale student assessment for improving student success". *Canadian Journal of Education*, vol. 29, núm. 3, págs. 873 a 883.
- Unión Europea. 2014. *Learning Assessments*. Documento de debate de la Unidad B4 para la Educación de la Dirección General de Cooperación Internacional y Desarrollo (DEVCO). Bruselas: Unión Europea.
- Unterhalter, E. 2019. "The Many Meanings of Quality Education: Politics of Targets and Indicators in SDG 4". *Global Policy*, vol. 10, págs. 39 a 51.
- Unterhalter, E. y Brighouse, H. 2007. "Distribution of what for social justice in education? The case of Education for All by 2015". M. Walker y E. Unterhalter (eds.). *Amartya Sen's capability approach and social justice in education* Palgrave Macmillan: Nueva York, págs. 67 a 86.
- USAID. 2011. *Education Opportunity Through Learning. USAID Education Strategy*. Washington D. C.: USAID.
- Verger, A. y Parcerisa, L. 2017. "Accountability and education in the post-2015 scenario: International trends, enactment dynamics and socio-educational effects". Documento de antecedentes para el *Informe de Seguimiento de la Educación en el Mundo 2017/2018*.
- Verger, A., Parcerisa, L. y Fontdevila, C. 2018. "The growth and spread of large-scale assessments and test-based accountabilities: a political sociology of global education reforms". *Educational Review*, vol. 71, núm. 1, págs. 5 a 30.
- Verger, A., Fontdevila, C. y Parcerisa, L. 2019. "Reforming governance through policy instruments: how and to what extent standards, tests and accountability in education spread worldwide". *Discourse: Studies in the Cultural Politics of Education*, págs. 1 a 23.
- Wagner, D.A. 2011. *Smaller, Quicker, Cheaper Improving Learning Assessments for Developing Countries*. París: IIPE.
- Wagner, D., Lockheed, M., Mullis, I., Martin, M., Kanjee, A., Gove, A. y Dowd, A. 2012. "The debate on learning assessments in developing countries". *Compare: A journal of comparative and international education*, vol. 42, núm. 3, págs. 509 a 545.
- Wagner, D.A., Wolf, S. y Boruch, R.F. (eds.) 2018. *Learning at the bottom of the pyramid: Science, measurement, and policy in low-income countries*. París: IIPE.
- Waldow, F., Takayama, K. y Sung, Y.K. 2014. "Rethinking the Pattern of External Policy Referencing: Media Discourses Over the 'Asian Tigers'. PISA Success in Australia, Germany and South Korea". *Comparative Education*, vol. 50, núm. 3, págs. 302 a 321.
- Walter, S. y Benson, C. 2012. "Language policy and medium of instruction in formal education". B. Spolsky (ed.). *The Cambridge Handbook of Language Policy*. Cambridge: Cambridge University Press, págs. 278 a 300.
- Weiss, I., Knapp, M., Hollweg, K. y Burrill, G. 2002. *Investigating the influence of Standards: A framework for research in mathematics, science and technology*. Washington D. C.: National Academy Press.
- Winthrop, R. y Anderson, K. 2013. "Can international large-scale assessments inform a global learning goal? Insights from the Learning Metrics Task Force". *Research in Comparative and International Education*, vol. 8, núm. 3, págs. 279 a 295.

Yasukawa, K., Hamilton, M. y Evans, J. 2017. "A comparative analysis of national media responses to the OECD Survey of Adult Skills: policy making from the global to the local?" *Compare: a journal of comparative and international education*, vol. 47, núm. 2, págs. 271 a 285.

Yasunaga, M. 2014. *Non-formal education as a means to meet learning needs of out-of-school children and adolescents*. Publicación en línea: UNESCO/UNICEF.

Zadja, J. 2014. *Understanding the quality debate in education*. Documento de sesión. Facultad de Educación y Artes de la Universidad Católica de Australia (campus de Melbourne).

Anexo 1: Evaluaciones transnacionales del aprendizaje

ESTUDIOS INTERNACIONALES						
Evaluación	Ámbitos de la prueba	Grupos destinatarios	Organismo de realización	Años de realización	Lista actualizada de los países participantes	Estudios y análisis consolidados organizados por los organismos de realización
TIMSS TIMSS sobre aritmética TIMSS avanzado	Matemáticas, ciencia	4º y 8º curso De 4º a 6º curso Alumnos en el último año de la enseñanza secundaria superior	Asociación Internacional de Evaluación del Rendimiento Escolar (IEA)	1995, 1999, 2003, 2008, 2007, 2011, 2015 Próximo estudio: 2019	http://timss2015.org/timss-2015/about-timss-2015/ http://timssandpirls.bc.edu/timss2015/international-results/advanced/timss-advanced-2015/about-timss-advanced/	https://timssandpirls.bc.edu/isc/publications.html https://iccs.iea.nl/resources/publications.html https://icils.acer.org/ http://pub.iea.nl/national_reports.html
PIRLS PIRLS sobre alfabetización	Competencia lectora	4º curso De 4º a 6º curso		2001, 2006, 2011, 2016 Próximo estudio: 2021	http://timssandpirls.bc.edu/pirls2016/international-results/pirls/about-pirls-2016/pirls-2016-countries/	
ICCS	Educación cívica y ciudadanía	8º curso		2009, 2016	https://iccs.iea.nl/cycles/2016/countries.html	
ICILS	Conocimientos informáticos y de la información	8º curso		2013 Próximo estudio: 2018	https://www.iea.nl/icils	
LaNA	Lectura y aritmética	De 4º a 6º curso		En fase de prueba	<i>Destinado a los países en desarrollo; no está listo para incorporar a los participantes en el TIMSS o el PIRLS</i>	

ESTUDIOS INTERNACIONALES						
Evaluación	Ámbitos de la prueba	Grupos destinatarios	Organismo de realización	Años de realización	Lista actualizada de los países participantes	Estudios y análisis consolidados organizados por los organismos de realización
PISA	Lectura, matemáticas y ciencia Ámbitos innovadores Conocimientos financieros (opcional)	Alumnos de 15 años	Organización de Cooperación y Desarrollo Económico (OCDE)	2000, 2003, 2006, 2009, 2012, 2015, 2018	https://www.oecd.org/pisa/aboutpisa/pisa-2018-participants.htm	http://www.oecd.org/pisa/publications/ https://www.oecd.org/pisa/pisa-for-development/
PISA-D	Lectura, matemáticas y ciencia	Alumnos de 15 años		2017	https://www.oecd.org/pisa/aboutpisa/pisa-for-development-participating-countries.htm	

ESTUDIOS REGIONALES						
Evaluación	Ámbitos de la prueba	Grupos destinatarios	Organismo de realización	Años de realización	Lista actualizada de los países participantes	Estudios y análisis consolidados organizados por los organismos de realización
PASEC ⁴⁸	Matemáticas y competencia lectora	2º y 6º curso	Conférence des ministres de l'Éducation des États et gouvernements de la Francophonie (Conferencia de Ministros de Educación de los Países de Habla Francesa) (CONFEMEN)	2014 Próximo estudio: 2019	http://www.pasec.confemen.org/wp-content/uploads/2015/12/Resume_Pasec2014_FR_BD1.pdf (véase el apartado "Contexte, objectifs et méthodologie de l'enquête PASEC2014")	http://www.pasec.confemen.org/pasec2014-3/
PILNA	Alfabetización y aritmética	4º y 6º curso	Programa de Evaluación y de Calidad Educativa de la Comunidad del Pacífico (SPC)	2012 y 2015	https://www.eqap.org.fj/getattachment/Resourses/Final-Regional-Report.pdf.aspx (véase el apartado "Resumen")	https://www.eqap.org.fj/work/Assessment.aspx
PERCE	Matemáticas, ciencia y lectura	3er y 6º curso	Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE)	1997	https://unesdoc.unesco.org/ark:/48223/pf0000243532 (véase el apartado "Introducción general")	http://www.unesco.org/new/es/santiago/education/education-assessment-llece/first-regional-comparative-and-explanatory-study/
SERCE				2006		http://www.unesco.org/new/es/santiago/education/education-assessment-llece/second-regional-comparative-and-explanatory-study-serce/
TERCE				2013		http://www.unesco.org/new/es/santiago/education/education-assessment-llece/terce/documents/

48 Antes de 2014, el PASEC llevó a cabo 35 evaluaciones nacionales en más de 20 países francófonos de África, Asia y el Oriente Medio.

ESTUDIOS REGIONALES						
Evaluación	Ámbitos de la prueba	Grupos destinatarios	Organismo de realización	Años de realización	Lista actualizada de los países participantes	Estudios y análisis consolidados organizados por los organismos de realización
SACMEQ o SEACMEQ	Lectura (1995-1999) Lectura y matemáticas (resto de los ciclos de realización) Conocimientos sobre el VIH/sida en 2006-2010	6º curso	Consorcio de África Meridional y Oriental para la Supervisión de la Calidad de la Educación (SACMEQ/SEACMEQ)	1995-1999, 2000-2004, 2006-2010, 2012-2014	http://www.sacmeq.org/?q=sacmeq-projects/sacmeq-iv	http://www.sacmeq.org/?q=publications
SEA-PLM	Lectura, escritura, matemáticas y ciudadanía global	5º curso	Organización de los Ministros de Educación de Asia Sudoriental, en colaboración con el UNICEF y el Consejo Australiano de Investigación Educativa (ACER)	2018-2019	http://www.seaplml.org/seaplml/index.php/about/countries	http://www.seaplml.org/seaplml/index.php/capacity-building-2

Nota: A los efectos del presente estudio, las EGRA y las EGMA (véase el anexo 2) *no* se consideran evaluaciones transnacionales porque, si bien los modelos de evaluación proporcionados en distintos países se basan en el mismo marco metodológico, su propósito *no* es facilitar las comparaciones entre los países.

Anexo 2: Evaluaciones nacionales con marcos metodológicos comunes

Evaluaciones de la lectura en los primeros cursos (EGRA) y evaluaciones de matemáticas en los primeros cursos (EGMA). Son evaluaciones orales, realizadas en persona y, por lo general, en la escuela. Sus resultados se emplean con frecuencia para promover reformas educativas. Están diseñadas para ofrecer medidas sencillas y económicas en cuestión de lectura y aritmética. Son obra del Research Triangle Institute (RTI International), que concibió la EGRA en 2006 y la EGMA, unos años después. La EGRA (lectura) evalúa el reconocimiento de las letras, la conciencia fonémica, la lectura de palabras sencillas y la comprensión oral. La EGMA (matemáticas) evalúa el reconocimiento numérico, las comparaciones y la ordenación de conjuntos de objetos. Más de 30 organizaciones en 70 países han utilizado la EGRA, que está adaptada a más de 120 idiomas. La EGMA, por su parte, se ha llevado a cabo en 22 países y está adaptada a 24 lenguas (Beggs, 2016).

Evaluaciones impulsadas por los ciudadanos. Suelen ser evaluaciones impulsadas por ciudadanos u organizaciones de la sociedad civil, en lugar de por los gobiernos. Se realizan en los hogares, en vez de en la escuela, y valoran competencias básicas de lectura y aritmética. Las evaluaciones impulsadas por los ciudadanos pueden proporcionar estimaciones recurrentes de los niveles educativos básicos de los niños y (hasta la fecha) tienden a presentar un diseño similar y a realizarse de una forma parecida. Catorce países están poniendo en marcha evaluaciones lideradas por los ciudadanos, que organizan a través de la Red de Acción Popular para el Aprendizaje (PAL Network). El movimiento se originó en la India (2005), a la que siguieron el Pakistán (2006), Kenya y la República Unida de Tanzania (2007), Uganda (2009), Malí (2010), el Senegal (2010), México (2013), Nigeria (2014), Bangladesh, el Camerún, Ghana y Mozambique (2016), y Nepal (2017). En conjunto han sometido a pruebas a 7,5 millones de niños en situaciones de extrema vulnerabilidad por medio de 48 evaluaciones llevadas a cabo en 30 idiomas, y han contado con la ayuda de 600.000 voluntarios (PAL Network, 2019).



La promesa de las evaluaciones de aprendizaje a gran escala

Reconocer los límites para generar oportunidades

Las iniciativas educativas en el marco de la Agenda 2030 para el Desarrollo Sostenible otorgan una importancia renovada a la instauración de un aprendizaje igualitario, eficaz y pertinente para todos. Así pues, se están realizando mayores esfuerzos de seguimiento del progreso hacia ese objetivo. Las evaluaciones de aprendizaje a gran escala (LSLA), que han adquirido importancia como componente estratégico clave para realizar un seguimiento de los resultados de los procesos educativos, son fundamentales para lograrlo. De hecho, las LSLA se han propagado por todo el mundo en los dos últimos decenios y, en la actualidad, se utilizan en más de la mitad de los países de todo el mundo. La utilidad de las LSLA como factores impulsores del progreso educativo está ampliamente documentada en una creciente bibliografía especializada.

Sin embargo, mientras los posibles usos de esta herramienta de seguimiento siguen creciendo, se sabe menos sobre sus límites. Las consecuencias imprevistas de las LSLA han suscitado preocupación entre diversas partes interesadas e investigadores en materia de educación. Sus posibles efectos secundarios se derivan del modo en que están diseñadas y del posible uso inadecuado de los datos de evaluación para impulsar políticas y prácticas educativas. Lo ideal sería que hubiese un entendimiento equilibrado de cómo pueden utilizarse para promover la agenda educativa y de las posibles dificultades que se generan cuando se usan de manera indebida. Esta publicación contribuye al debate mundial en curso examinando la información con la que contamos sobre los aspectos positivos de las LSLA y aclarando al mismo tiempo las consecuencias imprevistas que podrían conllevar.

