



EVALUACIÓN AL SERVICIO DE LOS APRENDIZAJES. Tendencias Globales. Desafíos Regionales

Al cabo de una etapa de instalación de los sistemas nacionales de evaluación de aprendizajes muchos países entraron en una fase de revisión de lo realizado para luego explorar nuevas alternativas. Después de más de una década de continuidad y cambios se procedió a examinar al menos tres aspectos fundamentales de las políticas adoptadas:

- La configuración de los sistemas de evaluación en la política educativa;
- La necesidad de mejorar la calidad técnica de los diversos componentes de los sistemas de evaluación; y
- Las estrategias de uso y difusión de los resultados de las evaluaciones.

Marco institucional de las agencias o institutos de evaluación.

Los países han ensayado diferentes marcos institucionales para la administración de sus sistemas de evaluación educativa y explorado las mejores formas de atender los intereses y metas de políticas de Estado en educación. En general, a pesar de los cambios y, en algunos casos, la inestabilidad política, ha predominado una tendencia a sostener los sistemas nacionales de evaluación, y a reinstaurarlos o recomponerlos cuando sus actividades fueron discontinuadas en algún momento. En varios países, los arreglos institucionales más estables resultaron ser los que se instalaron fuera de la estructura orgánica de los ministerios de educación. Estos sistemas, generalmente denominados “institutos”, poseen mayor autonomía administrativa y técnica que otros órganos de línea y suelen llevar a cabo sus tareas de medición y devolución de resultados con más agilidad y consistencia que los sistemas dependientes de los ministerios.

Resumen

Las políticas de evaluación de logros de aprendizaje en América Latina datan de alrededor de tres décadas atrás aunque la mayor parte de los sistemas nacionales de medición de calidad se institucionalizaron durante los años noventa. Algunos países, como Argentina, Chile, Brasil y Colombia vienen realizando desde hace muchos años aplicaciones continuas de pruebas de rendimiento académico y las unidades de medición a cargo gozan de altos grados de legitimidad y continuidad técnica a pesar de la rotación de cuadros técnicos que se producen por cambios de gobierno nacional. Otros sistemas, como los de Ecuador, Guatemala, Bolivia, Perú y México han sido más vulnerables a los cambios políticos, lo cual ha determinado que las informaciones del sistema no siempre sean públicamente difundidas, o que se produzcan recambios en los equipos técnicos con el consecuente impacto sobre la calidad de los instrumentos, operativos, o del diseño e implementación de estrategias apropiadas para la difusión de los datos. Se resumen en este estado del arte documentos producidos en el marco de las actividades del Grupo de Trabajo sobre Estándares y Evaluación del PREAL disponibles en línea en:

<http://www.grade.org.pe/gtee-preal>

<http://www.thedialogue.org>

<http://campus-oei.org>

MÉXICO, luego de haber contado con un sistema de medición dependiente de la Secretaría de Educación Pública durante muchos años, creó el Instituto Nacional de Evaluación de la Educación (INEE), primero como organismo semiautónomo y, desde 2013 con una institucionalidad autónoma con mayores atribuciones para la toma de decisiones técnicas y administrativas sobre los operativos de evaluación y la coordinación del sistema educativo nacional. También Argentina intentó en el año 2001 la creación de un instituto semiautónomo de evaluación, el IDECE, aunque inesperados cambios de gobierno nacional en ese mismo año determinaron su disolución y la reinstalación del sistema nacional dentro del organigrama regular del Ministerio de Educación.



Dos ejemplos de este tipo de institucionalidad son el ICFES en Colombia y el INEP en Brasil. En Colombia, si bien existe una unidad de evaluación dentro del Ministerio de Educación Nacional, el protagonismo técnico y operativo en la realización de diferentes evaluaciones nacionales lo ha tenido el ICFES- Instituto Colombiano de Fomento a la Educación Superior. Este organismo, sin ser totalmente independiente del Ministerio, ya que este último lo financia y preside, goza de amplia discreción y autonomía en la toma de decisiones técnicas con respecto a algunas de las pruebas nacionales. En Brasil, las evaluaciones nacionales de aprendizajes las administra INEP- Instituto Nacional de Estudios e Investigaciones Educativas, que también opera independientemente del Ministerio de Educación. En ambos casos, los operativos de evaluación se realizan periódicamente y los equipos técnicos han podido consolidarse y desarrollar nuevas tecnologías para la medición de los aprendizajes. Otro factor que parece contribuir a la estabilidad y permanencia de los sistemas de medición de la calidad de la educación y los logros de aprendizaje es la existencia de leyes nacionales que ordenen su creación y sostenimiento. Un ejemplo es el sistema de Pruebas Nacionales en República Dominicana, que si bien funciona en condiciones organizacionales relativamente precarias, los operativos anuales de evaluación siguen realizándose regularmente, tal como indica la Ley General de Educación y un conjunto de ordenanzas nacionales específicas. Algo similar a lo que ocurrió en Perú, por ejemplo, donde la nueva Ley General de Educación prescribe la creación y sostenimiento permanente de un Instituto Peruano de Evaluación, Acreditación y Certificación Educativa.

Los sistemas estatales de evaluación de resultados de aprendizajes, por su parte, se administran enteramente desde las secretarías estatales o distritales de educación, y este arreglo institucional parece funcionar adecuadamente en varios contextos. El caso del SARESP, Sao Paulo, muestra un arreglo institucional diferente, donde el protagonismo técnico lo tiene una institución independiente del gobierno. Por la escala de estos sistemas educativos, más pequeña y más fácilmente administrable, los sistemas de evaluación logran mayores grados de comunicación y trabajo orgánico con otras oficinas sectoriales, facilitando así el flujo de datos y magnificando las posibilidades de impacto en la toma de decisiones de política educativa -fenómeno que, en muchos países, no se da en los niveles centrales de la administración del sector. Son sistemas, también, que logran mayores grados de participación de los actores locales, especialmente docentes, padres de familia y organizaciones académicas vinculadas a la formación docente.

LAS AGENCIAS AUTÓNOMAS, O INSTITUTOS, si bien pueden tener mayores grados de autonomía funcional y legitimidad técnica, también pueden presentar problemas. El mayor riesgo es que se desvinculen de las necesidades de información de los Ministerios, convirtiéndose en programas de medición de alta calidad técnica de escaso impacto en la toma de decisiones y el diseño e implementación de políticas para la mejora de la calidad y equidad educativa. En otros casos sucede lo inverso. Sistemas anteriormente administrados por instituciones independientes se encuentran hoy administrados desde los ministerios de educación. Tal es el caso de Chile, donde la organización e implementación inicial del SIMCE, hacia fines de los años 80, fue delegada por el Ministerio de Educación en universidades suficiente capacidad técnica para llevar a cabo estas tareas. Actualmente, el sistema de evaluación se administra enteramente desde la Agencia de la Calidad de la Educación, entidad creada por Ley de la República que forma parte de la nueva institucionalidad del sistema nacional de educación desde fines de los años 90. No todos los casos de transferencia de responsabilidades a los ministerios de educación resultan exitosos o convenientes. En Ecuador, por ejemplo, el término del programa de evaluación APRENDO, financiado enteramente con fondos de cooperación internacional, significó el final de los operativos. Si bien se planificó la transición del sistema a la administración pública ministerial, la ausencia de recursos impidió la continuidad de las evaluaciones tal como venían realizándose y no se diseñaron nuevos programas de evaluación. Este fenómeno - común a casi todos los países que iniciaron sus sistemas de evaluación con financiamiento y apoyo técnico de los organismos de cooperación internacional, principalmente el BID y el Banco Mundial, tiene excepciones como Uruguay y Argentina, que hoy financian con fondos del presupuesto ordinario sus sistemas de medición dentro de la estructura orgánica ministerial y la continuidad de los operativos no se ha visto mayormente amenazada.



Medición de aprendizajes, desarrollo curricular y formulación de estándares.

Una de las limitaciones de quienes a su cargo el diseño de instrumentos de medición son los currículos nacionales que ofrecen definiciones poco claras, y escasamente operacionales, de lo que se espera que los alumnos puedan realizar con los conocimientos contenidos en ellos. Casi siempre la matriz curricular para el desarrollo de los instrumentos de evaluación se realiza mediante el diseño de una “tabla de especificaciones” que permite construir los ítems de las pruebas cubriendo un número determinado de objetivos prioritarios, aunque éstos no representen la totalidad o complejidad del currículo oficial. Lo negativo en esta situación es que el marco de interpretación de los resultados es débil puesto que no refleja acuerdos, profesionales y políticos, sobre expectativas de aprendizaje. Esto es, que deben saber y saber hacer los estudiantes del sistema. A lo que se agrega el problema de priorizar contenidos que pueden ser medidos más fácilmente, pero no son los más importantes para los actores del sistema. Los problemas se agravan por la débil comunicación que frecuentemente existe entre las unidades de evaluación y las dedicadas al diseño y desarrollo curricular a nivel nacional, por razones que incluyen, entre otras, el diferente origen y formación profesional de sus técnicos. Esto, junto a la poca claridad de las metas de aprendizaje, suele ser uno de los mayores obstáculos para que las actividades de evaluación sean coherentes con esas metas y puedan efectivamente facilitar su consecución.

En algunos países se está avanzando en la dirección de establecer algunas especificaciones, consensuadas entre los evaluadores, sobre expectativas de aprendizaje, aun cuando los procesos de desarrollo curricular del país no estén alineados con ellas. La evidencia indica que, actualmente, muchos países latinoamericanos cuentan con equipos de profesionales de la evaluación dedicados a la especificación de parámetros y metas curriculares que permitan diseñar pruebas más enfocadas y avanzar en la formulación de estándares de aprendizaje y desempeño alineados con las evaluaciones. A falta de resoluciones a favor de la formulación de estándares el trabajo de las unidades de medición se constituye en una instancia de avance importante en esa dirección. Casos que ilustran esta tendencia son los Uruguay y Colombia, así como los sistemas sub nacionales de medición como de Aguascalientes en México y Bogotá, en Colombia.

VALIDACIÓN DE INSTRUMENTOS

Uno de los primeros pasos de validación técnica durante el diseño de los sistemas de evaluación estandarizada consiste en la toma de decisiones sobre el tipo de cobertura poblacional y curricular de los instrumentos de medición. Tales decisiones, que refieren principalmente a las poblaciones y a los contenidos de aprendizaje evaluados, deben estar ligadas coherentemente al tipo de uso que se hará posteriormente de los resultados. Con respecto a la cobertura poblacional, es necesario distinguir entre países que aplican pruebas muestrales y los que realizan mediciones censales. Ambos tipos de pruebas sirven propósitos diferentes. Las pruebas *censales*, aplicadas al universo de estudiantes en ciertos niveles de escolaridad, arrojan resultados por establecimiento, o por alumnos, que permite establecer sistemas de incentivos o sanciones de acuerdo a los logros obtenidos (altas implicancias), o bien pueden facilitar la focalización de apoyos centrales a aquellos centros educativos que presentan mayores necesidades y dificultades de rendimiento académico. Las aplicaciones *m muestrales* apuntan a recabar informaciones sobre diferentes grupos o poblaciones del sistema, para poder detectar problemas generales de aprendizaje y diseñar estrategias integrales de apoyo a todas las escuelas y estudiantes de un determinado grupo social, o una determinada región del país, generalmente las que más lo necesitan.

En algunos países se está avanzando en la dirección de establecer algunas especificaciones, consensuadas entre los evaluadores, sobre expectativas de aprendizaje, aun cuando los procesos de desarrollo curricular del país no estén alineados con ellas. La evidencia indica que, actualmente, muchos países latinoamericanos cuentan con equipos de profesionales de la evaluación dedicados a la especificación de parámetros y metas curriculares que permitan diseñar pruebas más enfocadas y avanzar en la formulación de estándares de aprendizaje y desempeño alineados con las evaluaciones. A falta de resoluciones a favor de la formulación de estándares el trabajo de las unidades de medición se constituye en una instancia de avance importante en esa dirección. Casos que ilustran esta tendencia son los Uruguay y Colombia, así como los sistemas sub nacionales de medición como de Aguascalientes en México y Bogotá, en Colombia.



Tipos de instrumentos y cobertura general de las evaluaciones.

La mayor parte de los sistemas nacionales de evaluación nacional miden resultados en las áreas curriculares de matemáticas, lenguaje y comunicación, ciencias sociales y ciencias naturales. En algunos casos se evalúan logros en lenguas extranjeras, especialmente en inglés y francés. En países multilingües como Perú o Bolivia, se aplican pruebas traducidas en lenguas vernáculas. En educación media es común que el área de ciencias naturales sea evaluada diferencialmente en biología, física y química, mientras que el área de ciencias sociales se evalúa con pruebas diferentes para historia, geografía y educación cívica. Casi todos los países que tienen sistemas nacionales de evaluación miden logros de aprendizaje en educación primaria y secundaria. En ambos casos, las pruebas tienden a ser aplicadas al término de ciclos educativos de dos o tres años, bajo la lógica de que algunos aprendizajes no pueden lograrse al cabo de un año académico sino luego de períodos más largos que coinciden con ciclos de edad y desarrollo cognitivo, o de secuencias curriculares multianual. Chile, República Dominicana y Costa Rica, aplican pruebas “de salida” del nivel de educación media o secundaria (algunas de ellas denominadas “pruebas de Bachillerato”), como requisito de graduación o como criterio de selección a la educación superior. México, aplica este tipo de pruebas también como instancia de graduación de la educación primaria y acceso a la educación media superior, aunque el uso que se da a los datos es generalmente diagnóstico y no selectivo.

La cobertura y análisis de logro curricular que se pretenda realizar determinará si las pruebas estarán referidas a normas o referidas a criterios. La selección de uno u otro tipo de pruebas se realiza de acuerdo a los fines y usos que se pretende dar a los resultados. En cuanto al formato de las pruebas, todos los países utilizan preguntas cerradas de opción múltiple, y varios de ellos complementan los instrumentos con preguntas abiertas para la producción de textos, especialmente en el área de lenguaje y comunicación. La corrección de respuestas abiertas ha presentado problemas técnicos para algunos países al momento de definir o utilizar criterios homogéneos de evaluación entre correctores. En otros países el proceso resultó más exitoso y las correcciones arrojan datos pedagógicamente muy útiles para los docentes.

- **Las pruebas referidas a normas** son las que buscan comparar los logros entre diferentes grupos de estudiantes, sin pretender un análisis muy profundo del tipo y profundidad de conocimientos adquiridos por ellos.
- **Las pruebas referidas a criterios**, suponen el establecimiento de acuerdos sobre cuáles competencias o capacidades académicas básicas debieran lograr todos los estudiantes, luego definen una línea de corte o marca de logro mínimo o suficiente, y miden si los alumnos del sistema educativo alcanzan o no esa marca, y en qué medida lo hacen, como resultado de su experiencia escolar.
- **Una mayor cobertura puede restringir la capacidad de los instrumentos para recabar datos de aprendizaje con mayor profundidad.** Los sistemas nacionales que aplican pruebas referidas a criterios ofrecen, en general, fundamentaciones conceptuales y descripciones operacionales más sólidas sobre los aprendizajes y niveles de desempeño esperados. Las pruebas referidas a normas, por el contrario, se diseñan mediante un modelo psicométrico que selecciona los ítems de evaluación de acuerdo al índice de discriminación en los pilotajes de campo, pero que puede no reflejar la relevancia de los contenidos evaluados ni asegura que éstos representen el universo curricular de referencia.

Aunque actualmente muchos países están recurriendo más al uso de respuestas abiertas, pese a su mayor costo y dificultad de revisión y codificación, por ser de mayor utilidad para la identificación de los procesos correctos o equívocos que siguen los alumnos en sus esfuerzos por responder a las preguntas y problemas que les plantean las pruebas y para desarrollar propuestas tanto para el mejoramiento de la enseñanza como el aprendizaje de ciertos contenidos y el desarrollo de competencias.



Factores asociados al rendimiento académico.

Las pruebas de logro académico generalmente se acompañan con la aplicación de cuestionarios a diversos actores de la comunidad educativa, normalmente alumnos, docentes, directores y padres de familia, para recoger información sobre los factores escolares y extraescolares que pueden asociarse al desempeño de los alumnos. Los factores asociados al sistema escolar suelen ser indicadores sobre las condiciones de infraestructura, equipamiento didáctico, recursos profesionales del establecimiento escolar, actitudes de los alumnos hacia las disciplinas evaluadas, tipos de prácticas pedagógicas, y grados de implementación del currículo oficial en aula. Los factores extraescolares, por su parte, dicen relación con el nivel económico y sociocultural de los alumnos, según la información que las familias aportan sobre sus ingresos, grados de escolaridad de los padres y hermanos del alumno, servicios básicos de energía y agua potable en el hogar, y ciertos recursos culturales y de información en el hogar, tales como TV, radio, libros, periódicos y revistas. El valor de la información contextual radica en que los sistemas de medición pueden reportar los resultados de las pruebas de acuerdo al nivel socioeconómico y cultural de los alumnos, y de esa manera hacer afirmaciones más justas sobre la calidad de las escuelas, de acuerdo al grado de vulnerabilidad social de los estudiantes. Dentro de este modelo de análisis, destacan los esfuerzos experimentales por medir “valor agregado” que han realizado algunos sistemas subnacionales, como la provincia de Buenos Aires en Argentina que permite, mediante pruebas de rendimiento aplicadas en dos momentos en el tiempo, medir el progreso en aprendizajes que una cohorte de estudiantes del mismo contexto socioeconómico logra respecto de sí misma al cabo de un lapso determinado. Lamentablemente, este operativo en Argentina ha sido discontinuado y no se han realizado publicaciones de los resultados obtenidos en las mediciones.

CASOS NACIONALES

En *Uruguay*, las pruebas aplicadas son referidas a criterios. La selección de los ítems se hace buscando los que reflejan tipos de capacidades que el sistema educativo considera deben ser desarrolladas antes del término de ciertos niveles de escolaridad. Una vez diseñados se establece una línea de corte, que en este caso es el 60% de respuestas correctas del total de los ítems. Al reportar resultados, se describe el porcentaje de alumnos que alcanzó esa línea y los que presentan niveles insuficientes de desempeño. En *Colombia*, se utiliza un tipo de pruebas centradas en el desarrollo de habilidades cognitivas complejas que reflejan conocimientos disciplinares más actualizados alineados con contenidos curriculares oficiales, estándares internacionales y concepciones formativas que ponen de relieve la necesidad de educar ciudadanos democráticos y participativos. Varios de los países que comenzaron aplicando pruebas referidas a normas han optado por realizar un informe de resultados según criterios definidos a posteriori. Tales son los casos de *Chile, Argentina, Costa Rica y Brasil*. En estos países, si bien el modelo psicométrico es normativo, se establecen criterios y líneas de corte para el reporte con la intención de informar mejor a los usuarios del sistema.

Adecuación lingüística.

Tres países aplican pruebas en lenguas maternas originarias. Perú lo hace en quechua y aymará, Bolivia en quechua, aymará y guaraní, mientras que Guatemala ha evaluado estudiantes en las cuatro lenguas mayas predominantes, que representan al 84% de la población guatemalteca. En este último caso, sin embargo, no fue posible materializar el operativo en su totalidad pues existe un alto grado de deserción escolar en escuelas de población maya-hablante.



Difusión de resultados y promoción de una cultura de la evaluación.

Tres grandes grupos de audiencias o destinatarios de los datos e informaciones producidos por los sistemas nacionales y sub-nacionales de medición de aprendizajes son claramente identificables. Un primer grupo, conformado por las autoridades políticas de gobierno, poderes ejecutivo y legislativo, la alta dirección del sector educativo de cada país. Un segundo grupo está integrado por los actores sociales, principalmente opinión pública, políticos, organizaciones sociales, académicas y no gubernamentales. El tercer grupo está integrado por los actores educativos locales, principalmente docentes, directores, supervisores de escuelas, alumnos, y padres y madres de familia.

A lo largo de su desarrollo, los sistemas nacionales de medición han utilizado formatos diferentes para reportar los datos obtenidos con la intención de hacerlos más útiles según tipos de destinatarios que se busca informar o las potenciales audiencias interesadas en usarlos.

En general, se observa una evolución hacia formas que apuntan crecientemente a la devolución de resultados en los niveles escolares, con particular énfasis en la presentación de datos curricular y pedagógicamente útiles. Esto contrasta con la tendencia inicial de algunos países, especialmente los que iniciaron sus sistemas de evaluación a principios de los años noventa, de producir informes generales que presentaban datos muy generales, en términos de porcentajes globales de rendimiento por asignatura y comparativos entre las diferentes subpoblaciones estudiantiles evaluadas, cuya utilidad estaba mayormente limitada a la selección de grupos o sectores en los cuales podría focalizarse algún tipo de intervención o, lo que ha sido más común, para movilizar a la opinión pública sobre el deficiente nivel de logros educacionales.

Informes para autoridades de gobierno y alta dirección pública, su finalidad es informar a las autoridades políticas y técnicas superiores para facilitar la toma de decisiones de política educativa con la base en indicadores relacionados con *procesos* y los *resultados* de la educación, que actúan como complemento de los indicadores tradicionales de *insumos* educativos tales como salarios, infraestructura o equipamiento.

Reportes a actores educativos locales. Los sistemas de medición de altas implicancias, por su responsabilidad de informar sistemáticamente a los usuarios directos de los datos, hacen públicos los resultados por escuela o por alumnos como el SIMCE en Chile o el ICFES en Colombia. En países donde las pruebas no tienen consecuencias para, los esfuerzos por divulgar públicamente los resultados han sido esporádicos y dependen de la voluntad y capacidad de los medios por hacer una cobertura periódica de los resultados.

Devolución con fines de mejoramiento curricular y pedagógico, ofrece información que permite a los reflexionar sobre el desempeño de los estudiantes y elaborar estrategias de mejoramiento ante dificultades de aprendizaje. SIMECAL de Bolivia, entrega resultados a las escuelas por medio de un informe institucional sobre el rendimiento de sus alumnos que incluye el puntaje promedio institucional, una descripción operacional de los niveles de logro definidos para cada área, y el porcentaje de alumnos de la escuela ubicados en cada nivel. La Secretaría de Educación de Bogotá entrega resultados por escuela. Los datos se reportan por asignatura, desagregados por competencia y niveles de desempeño alcanzados para cada una de ellas. En Brasil, los estados de Sao Paulo y Paraná, promueven la participación de las escuelas en la elaboración de los informes. Cada centro educativo se responsabiliza por el procesamiento primario de los datos usando formato y las pautas indicadas desde la unidad de evaluación de la Secretaría. Posteriormente, esos datos son elevados a la administración central donde son nuevamente procesados y verificados.

Informes específicos para padres de familia. En casi todos los países es el menos atendido por los sistemas de medición y evaluación. Países con pruebas censales solicitan a las escuelas que compartan la información con los padres de familia, pero en forma de recomendación, sin ningún tipo de prescripción de recomendaciones pedagógicas u organizacionales para elevar los logros de aprendizaje.



Uso de los resultados de las evaluaciones.

Constituye uno de los principales problemas en la evaluación de aprendizajes dependiendo del tipo de consecuencias que implique para los actores del sistema. Se espera que, cualesquiera sean las audiencias y las estrategias para llegar a ellas, los resultados de las evaluaciones dejen de ser simplemente **datos** para transformarse en **información** sobre los logros de aprendizaje de los alumnos. Esto significa que quienes tienen a su cargo la difusión deben asumir la responsabilidad de interpretarlos, reorganizarlos y presentarlos de una manera clara y diferenciada de acuerdo al uso que se supone diferentes audiencias dará a la información. Otro factor que afecta la posibilidad de impacto es el tiempo de devolución de los resultados, en términos de frecuencia y periodicidad de entrega. El uso que los actores educativos puedan dar a las informaciones se verá estimulado si los datos se difunden poco después de las aplicaciones y si las entregas se realizan periódicamente. Esto puede contribuir al afianzamiento del sistema de evaluación mientras se genera una demanda de datos en forma oportuna, es decir durante momentos clave para el diseño de políticas de mejora educativa, planificación de cobertura o entrega curricular en el ámbito escolar. Otros factores que inciden en el impacto son la participación de los actores en el proceso evaluativo; la calidad de los informes y la voluntad política por tomar decisiones informadas. Asimismo, es importante distinguir entre periodicidad de evaluación y entrega de resultados. Perú, por ejemplo, modificó su plan original de realizar operativos nacionales en varios grados y disciplinas cada dos años, ampliando el ciclo a tres. Chile ha modificado la periodicidad de sus pruebas, intercalando grados y materias de manera de permitir un mayor tiempo y recursos para el análisis de datos y el mejoramiento del sistema en general.

Uso para toma de decisiones o el diseño de programas de mejora.

En Chile, las evaluaciones SIMCE se usan, y han usado, para focalizar políticas de mejora escolar en al menos tres tipos de intervenciones:

- el uso de los resultados como uno de los dos indicadores más importantes en la selección de las escuelas que ingresaron a los programas focalizados de atención educativa;
- la utilización de los datos para la asignación de fondos concursables de proyectos de mejoramiento educativo en las escuelas; y

Los datos se transforman en información cuando comunican a los actores del sistema un mensaje intencional sobre los aprendizajes logrados, idealmente de acuerdo a un marco conceptual que permita interpretar los resultados en relación a un conjunto de criterios claros sobre los aprendizajes y desempeños esperados de los estudiantes.

- el uso de los datos como uno de los indicadores principales para otorgar incentivos docentes mediante el SNED- Sistema Nacional de Evaluación del Desempeño de los Establecimientos Educativos.

En México, un programa paradigmático de mejora escolar en escuelas urbano-marginales es el PEC- Programas de Escuelas de Calidad, hoy transformado en una política pública de extensión de la jornada escolar. Varios de los estados realizaron la selección de las escuelas participantes en función del contexto socioeconómico en el que laboran, así como de sus resultados en pruebas de aprendizajes nacionales o estatales. Las mismas pruebas son utilizadas, entre otros indicadores para monitorear el progreso educativo de las escuelas participantes. Las evaluaciones en Uruguay fueron utilizadas para fundamentar la ampliación de un programa de equidad para Escuelas de Tiempo Completo. Por el tipo de modelo de evaluación que se realiza en este país los datos sirvieron para demostrar que este tipo de escuelas logran mejores resultados que otras en contextos socioeconómicos semejantes. En el nivel subnacional, la toma de decisiones de política en función de los resultados de las evaluaciones es un fenómeno más extendido, especialmente para el diseño de programas de capacitación pedagógica y de gestión escolar. Destacan los sistemas de evaluación de Bogotá, Sao Paulo y Paraná, que han puesto en marcha programas integrales y continuos de capacitación docente y directivos escolares para elevar el rendimiento académico demostrado por los alumnos en las pruebas estandarizadas locales. En Bogotá, adicionalmente, se reporta que muchas de las decisiones sobre inversión en infraestructura escolar, programas alimentarios escolares y otros insumos se realizan teniendo en cuenta los resultados de las evaluaciones de aprendizaje. De la misma manera, los datos son utilizados para seleccionar y monitorear a las escuelas privadas que albergan matrícula pública mediante subvenciones per cápita.



Uso pedagógico en las escuelas.

Algunos sistemas de medición de aprendizajes ofrecen herramientas pedagógicas y asistencia técnica a las escuelas para el mejoramiento del desempeño de los estudiantes. Esto, mediante la elaboración de informes analíticos de los resultados de cada aplicación de pruebas para generar debates profesionales en torno a los resultados sobre diferentes ítems de las pruebas. De esta manera, los docentes pueden aplicar la prueba administrada centralmente y practicar individualmente el ejercicio de corrección. En Colombia, para los mismos propósitos, se han realizado talleres de capacitación en interpretación de resultados para 1400 escuelas. La Secretaría de Educación de Bogotá diseñó dos programas de intervención. Uno, denominado “Nivelación para la Excelencia”, mediante el cual se brinda apoyo en la gestión, infraestructura y en los recursos pedagógicos a las escuelas de peores resultados. Para implementar dispositivos de apoyo, especialmente en el componente pedagógico, se contratan universidades como agencias oferentes de la capacitación, quienes trabajan con los docentes y directores. La capacitación se realiza mediante visitas periódicas a las escuelas durante un año completo.

Uso de altas implicancias y responsabilización política.

Tanto Chile como México usan los resultados de las evaluaciones de rendimiento escolar y logros de aprendizaje como indicadores complementarios para la asignación de incentivos docentes. En el caso de Chile, se hace mediante el Sistema Nacional de Evaluación del Desempeño de los Establecimientos Educativos Subvencionados (SNED) que otorga un bono anual al equipo docente de escuelas que logran, entre otros desempeños profesionales, elevar el rendimiento académico de sus estudiantes respecto de mediciones anteriores. Los resultados del SIMCE también se utilizan para seleccionar y monitorear los logros de las escuelas que solicitan y obtienen financiamiento para Proyectos de Mejoramiento Educativo (PME) un sistema de fondos concursables para desarrollar y ejecutar propuestas de mejoramiento institucional. En México, los docentes pueden postular al programa de Carrera Magisterial, un sistema de incentivos salariales que también utiliza los resultados de las pruebas de aprendizaje nacionales como uno de los indicadores de desempeño profesional.

Las pruebas internacionales incluyen ítems de opción múltiple, preguntas abiertas y/o demostraciones prácticas. Entre los factores contextuales destacan las características familiares de los alumnos, infraestructura escolar, preparación de los docentes, y actitudes y valores de los alumnos con relación a esas áreas. Las pruebas son diseñadas y coordinadas por agencias internacionales de evaluación, con distintos grados de participación por parte de las unidades de medición de la calidad educativa de los países. Algunas de estas agencias son privadas e independientes, como la Asociación Internacional para la Evaluación del Logro Académico (IEA), o ETS (Servicios de Evaluación Educacional). Otras son agencias intergubernamentales donde participan las autoridades políticas o técnicas de administración y evaluación educativa. Ejemplos de estas últimas son las pruebas PISA de la OCDE (Organización para la Cooperación y el Desarrollo Económico), el Laboratorio Latinoamericano para la Evaluación de la Educación (LLECE) de UNESCO-OREALC. Las pruebas de la IEA o de la OCDE son de cobertura global, mientras que otras son de cobertura regional. Argentina, Bolivia, Brasil, Chile, Colombia, Cuba, Honduras, México, Paraguay, Uruguay, República Dominicana, Perú y Venezuela han participado en las pruebas y sus experiencia varía según las capacidades técnicas con que ingresaron al proyecto, y las condiciones políticas que acompañaron el proceso. La voluntad de hacer públicos los resultados y la estabilidad de las autoridades, políticas y técnica, inciden en el tipo de impacto que tienen los resultados sobre las decisiones de política. El impacto en América Latina ha sido débil, debido a la falta de estrategias de comunicación y utilización de los resultados, así como a la insuficiente definición de los propósitos y objetivos de las evaluaciones

A partir del 2013 se creó el sistema obligatorio de evaluación docente que evalúa la idoneidad de maestros y directores de centros escolares para su ingreso, permanencia y promoción en la carrera. En Brasil, el estado de Sao Paulo, los datos censales del SARESP constituyen un indicador determinante para acceder al ciclo siguiente de escolaridad básica. Un desempeño insuficiente en las pruebas del SARESP (menos de 50% de acierto) puede ser revisado por las autoridades si el rendimiento del estudiante ha sido muy bueno durante el ciclo y se presume que por alguna razón especial no pudo rendir al nivel de su potencial el día de la aplicación.



A partir del 2013 se creó el sistema obligatorio de evaluación docente que evalúa la idoneidad de maestros y directores de centros escolares para su ingreso, permanencia y promoción en la carrera. En Brasil, el estado de Sao Paulo, los datos censales del SARESP constituyen un indicador determinante para acceder al ciclo siguiente de escolaridad básica. Un desempeño insuficiente en las pruebas del SARESP (menos de 50% de acierto) puede ser revisado por las autoridades si el rendimiento del estudiante ha sido muy bueno durante el ciclo y se presume que por alguna razón especial no pudo rendir al nivel de su potencial el día de la aplicación. En ese caso, el alumno tiene una segunda oportunidad para presentarse a la prueba.

Frecuentemente se cita en los documentos oficiales de evaluación educativa el propósito de hacer de los resultados de las mediciones una medida de responsabilidad político-administrativa por los avances y logros del sistema. Pero, son pocos los casos de rendición de cuentas del Estado frente la sociedad y ante la comunidad educativa por los resultados del aprendizaje. Uno de los factores que inhiben la difusión de resultados de aprendizaje como indicador de desarrollo educativo radica en que los resultados se ven condicionados por factores externos que no pueden ser intervenidos por el sistema escolar y el hecho que los resultados obtenidos en un momento del tiempo reflejan resultados de un procesos que duran ciclos más extensos que un período presidencial o de gestión ministerial. Aunque la mayoría de los países no ponen los resultados de la evaluación a disposición del escrutinio público hay dos países que escapan a esta tendencia. Uno es Chile, donde los resultados publicados anualmente por el SIMCE suelen tener una alta repercusión política, que se manifiesta en forma de artículos académicos y de opinión que utilizan los resultados y las bases de datos de las evaluaciones para cuestionar los logros de la reforma educacional en términos de calidad y equidad. En Colombia, la Secretaría de Educación de Bogotá, ha hecho públicas sus metas para lograr que el conjunto de escuelas distritales llegara a los 180 puntos de rendimiento promedio (en una escala de 0 a 306), y con un rendimiento global mínimo de 100 puntos en cada centro educativo.

Pruebas internacionales de aprendizaje.

Las pruebas internacionales son instrumentos de medición estandarizados que se aplican en varios países simultáneamente, en grados o grupos de edad seleccionados, y recogen información sobre logros de aprendizaje en diferentes disciplinas, y sobre los factores asociados que influyen en el de los estudiantes.

CONCLUYENDO, las políticas y sistemas nacionales de evaluación, sus avances y desafíos en materia de medición y evaluación de la calidad de los aprendizajes en América Latina tienden a:

- 1 Una mayor apertura a explorar alternativas institucionales para la organización y administración de las unidades de medición de la calidad de la educación y logros de aprendizaje.** Predominan tipos de organización como las unidades dependientes de los Ministerios de Educación, los organismos autónomos e institutos semiautónomos y universidades contratadas por los gobiernos. No es posible afirmar que una forma de organización sea mejor que otras. La elección depende del contexto político y de las capacidades técnicas y financieras para la evaluación a gran escala en cada país. El avance radica en la experimentación de formas alternativas de gestión, lo que indica que existe una fuerte voluntad de sostener y perfeccionar los sistemas de medición, aun cuando el impacto en algunos países no haya sido de gran magnitud (Ravela, Arregui, et.al., 2006, 2017)
- 2 Se han hecho esfuerzos por establecer expectativas claras sobre el tipo de competencias académicas que han de desarrollar los estudiantes, y que serán posteriormente evaluadas.** La mayor parte de ellos ha estado en manos de los equipos de evaluación, quienes ante la falta de referentes curriculares claros (Ferrer, 1999; Ferrer, 2004) han optado por proponer marcos conceptuales e indicadores de logro más específicos que faciliten una interpretación sustantiva de los resultados de las mediciones. El aspecto más débil en este terreno es el de la escasa vinculación que aún existe entre los equipos de evaluación y los equipos de desarrollo curricular en las sedes ministeriales del sector.
- 3 Se constatan avances la determinación consensuada de los objetivos de los sistemas nacionales de medición.** Concretamente, a quiénes se evalúa, con qué fines, y qué tipos de pruebas se aplican para



lograrlos. En general, parece existir bastante coherencia entre las decisiones sobre poblaciones a ser evaluadas y tipos de decisiones de política y pedagógicas que se busca informar por medio de las evaluaciones. Por ejemplo, los países que aplican pruebas muestrales sólo persiguen entender algunas aspectos del rendimiento del sistema, para diseñar medidas generales de política y promover la rendición de cuentas. Los sistemas que aplican evaluaciones censales apuntan a dos tipos de objetivos principales: promover una cultura de la evaluación mediante sistemas de incentivos o mecanismos de acreditación, directamente a los actores locales del sistema, por una parte. Por otra, ofrecer insumos de información altamente desagregados sobre desempeño escolar para facilitar la planificación de estrategias puntuales de intervención curricular y pedagógica.

4 Las formas en que se divulgan y usan los datos de las evaluaciones no siempre resultan útiles a las audiencias a las que destinan. Es necesario aprender del esfuerzo que están haciendo algunos países por reportar los datos de manera tal que las escuelas puedan compararse con instituciones semejantes y así obtener una medida más justa sobre la eficiencia de su labor pedagógica.

5 Las estrategias de difusión y uso de resultados, tanto en el caso de las pruebas nacionales como internacionales aún es muy débil la relación entre las agencias evaluadoras y los usuarios del sistema. Faltan estrategias comunicacionales para involucrar a usuarios de la información en instancias de elaboración de las pruebas, preparación de informes y definición de estrategias de difusión. Los países cuyos sistemas de evaluación tienen altas implicancias cuentan con estrategias de difusión más sólidas y se esfuerzan más por entregar información en forma desagregada según actores, escuelas e inclusive, por cursos y estudiantes.

6 Con relación al uso de las evaluaciones, es necesario distinguir entre sistemas que apuntan a fortalecer la responsabilidad política y administrativa por los resultados de aprendizaje, los que promueven rendiciones de cuentas, institucional, profesional y estudiantil, y los que buscan un impacto pedagógico y curricular al nivel local. Los primeros han tenido menos éxito, pues la escasa participación de la sociedad civil en los procesos evaluativos no ha favorecido movimientos importantes en demanda por mayores niveles de calidad.

Aunque, en algunos países se aprecia un mayor impacto mayor de los resultados en la opinión pública. Países que intentan promover la responsabilidad institucional, profesional y estudiantil cuentan con marcos normativos y mecanismos de gestión sólidos que aseguran un uso extendido y regular de los resultados de las evaluaciones y formatos de entrega de resultados.

7 Los sistemas sub-nacionales y la participación en pruebas internacionales son campos que merecen mención especial sobre todo en países donde se ha descentralizado la gestión escolar y se dan iniciativas tendientes a desarrollar sistemas de medición que acomoden mejor las necesidades locales o regionales. Dado el potencial de esos sistemas, puede adelantarse que las iniciativas han logrado grados importantes de consolidación y legitimación frente a la comunidad educativa, y otros actores.

8 Varios países han participado en pruebas internacionales y, al menos, en una evaluación regional comparada. Los resultados de la participación han sido diferentes para cada país, dependiendo de las capacidades técnicas y los contextos políticos en que se dieron. Si bien la experiencia ha significado una excelente oportunidad de aprendizaje y capacitación, el impacto de los datos recogidos ha sido bajo. La difusión masiva de los resultados, vía prensa escrita principalmente, ha generado opiniones críticas debido a las grandes diferencias de rendimiento entre los países de América Latina y otros de mayor desarrollo relativo. Los resultados han circulado poco como para informar decisiones y promover cambios en las prácticas pedagógicas, curriculares y de gestión.



Criterios y categorías de análisis para evaluar la formulación de políticas de evaluación de resultados del aprendizaje

La evaluación	No aporta porque....
Marco institucional	Adecuación según marco político y capacidades técnicas: <ul style="list-style-type: none"> - Estabilidad - Financiamiento y autonomía administrativa - Recursos humanos - Autonomía y capacidad para la difusión de resultados - Transparencia
Currículo y estándares	<ul style="list-style-type: none"> - Disponibilidad, adecuación y utilización de marcos curriculares nacionales para el diseño de instrumentos de medición. - Desarrollo y validación de estándares en armonía con el currículo vigente.
Instrumentos	<ul style="list-style-type: none"> - Marco conceptual sólido y explícito para la elaboración de matrices de referencia. - Validación de matrices de referencia e instrumentos. - Tipos de ítems o reactivos. - Adecuación cultural y lingüística (EBI). - Cobertura muestral o censal coherente con objetivos de la evaluación. - Estudio de contexto escolar y extraescolar para análisis de factores asociados al rendimiento.
Reportes	<ul style="list-style-type: none"> - Coherencia entre tipos de reporte y usos esperados (currículum, pedagogía, focalización de apoyos, capacitación docente, selección de alumnos, etc.) - Adecuación de reportes a audiencias-objetivo (claridad, guías de interpretación, sensibilización, etc.). - NES y valor agregado.
Difusión y usos	<p>Devolución:</p> <ul style="list-style-type: none"> - Tiempos - Alcance - Periodicidad <p>Impacto:</p> <ul style="list-style-type: none"> - Uso escolar - Toma de decisiones de política/ diseño de programas - Responsabilización política <p>Altas implicancias:</p> <ul style="list-style-type: none"> - Escuelas - Docentes - Alumnos
Sistemas subnacionales	Diferencias principales con sistemas nacionales: <ul style="list-style-type: none"> - Estándares - Cobertura censal - Participación de actores locales - Tiempos y formas de devolución de resultados - Uso centrado en el mejoramiento pedagógico
Pruebas internacionales	<ul style="list-style-type: none"> - Desarrollo de capacidades técnicas - Difusión de resultados e impacto en la opinión pública - Usos específicos de los resultados

Fuente: Ferrer, 2006, 2015; Ravela, 2006; 2005; 2016.



REFERENCIAS BIBLIOGRÁFICAS

- Arregui, P. et.al. Sobre Estándares y Evaluaciones. PREAL Ediciones. Santiago, Chile. www.grade.org.pe/preal. PDF
- Ferrer, G. 2004. Las reformas curriculares de Perú, Colombia, Chile y Argentina: ¿Quién responde por los resultados? GRADE. Lima, Perú.
- Ferrer, G. y Arregui, P. 2003. Las pruebas internacionales de aprendizaje en América Latina y su impacto en la calidad de la educación: Criterios para guiar futuras aplicaciones. PREAL Documento #2
- Ferrer, G. 1999. Aspectos del currículo prescrito en América Latina: Revisión de tendencias contemporáneas en currículo, indicadores de logro, estándares y otros instrumentos. Mimeo. <http://www.grade.org.pe/gtee-preal/docs/Ferrer.pdf>
- Guevara, G. y Backhoff, E. (Coord.). 2015. Las transformaciones del sistema educativo en México, 2013-2018, Fondo de Cultura Económica, INEE, México.
- INEE. Informe 2018. La Educación Obligatoria en México. INEE, México, D.F. www.inee.edu.mx
- Martínez Rizo, F., et.al. Informe Anual 2008 ¿Avanza o retrocede la calidad educativa? Tendencias y Perspectivas de la Educación Básica en México. 2008. INEE. México, D.F.
- Ravela, P. (Ed.) 2001. Los próximos pasos: ¿Hacia dónde y cómo avanzar en la evaluación de aprendizajes en América Latina? PREAL, Documento N° 20.
- Ravela, P. 2001. ¿Cómo presentan sus resultados los sistemas nacionales de evaluación educativa en América Latina? PREAL.
- Ravela, Arregui, Valverde, Wolfe, Ferrer, Martínez Rizo. 2008. Las evaluaciones educativas que América Latina necesita. PREAL Documentos No. 40. Santiago, Chile. www.grade.org.pe/upload/publicaciones/
- Rojas, C. y Esquivel J.M. 1998. Los sistemas de medición del logro académico en Latinoamérica. Banco Mundial, LCSHD Paper Series.
- Wolff, L. (1998) Evaluación educacional en América Latina: Progresos actuales y desafíos futuros. PREAL Ed. Washington,