

Edición 2026



Evaluaciones de Aprendizaje a Gran Escala en América Latina

De la medición a la transformación de los sistemas educativos

Organización: **David Saad**, Diretor-Presidente de Instituto Natura para América Latina

ÍNDICE

03 **PRESENTACIÓN**

PARTE I – EL PAPEL DE LAS EVALUACIONES

04 **Capítulo 1** – Por qué importa medir el aprendizaje

12 **Capítulo 2** – ¿Sirve cualquier evaluación de aprendizaje a escala?

PARTE II – SISTEMAS NACIONALES EN PERSPECTIVA

18 **Capítulo 3** – **Brasil:** solidez y complementariedad

23 **Capítulo 4** – **Argentina:** tradición y desafíos de consolidación

28 **Capítulo 5** – **Perú:** institucionalización con tensiones metodológicas

33 **Capítulo 6** – **Chile:** madurez, crisis y reinención

37 **Capítulo 7** – **Colombia:** sistema estructurado y consistente

42 **Capítulo 8** – **México:** de la referencia a la ausencia de evaluación nacional

48 **Capítulo 9** – Un cuadro comparado para América Latina

53 **Conclusión** – El futuro de las evaluaciones a gran escala



INTRODUCCIÓN

Este e-book reúne, organiza y profundiza una serie de artículos dedicados al papel de las evaluaciones de aprendizaje a gran escala en los sistemas educativos de América Latina.

Los textos, escritos en colaboración con Karina Stocovaz, Superintendente del Instituto Natura para los países de habla hispana, parten de una premisa central: en contextos marcados por profundas desigualdades educativas, no es posible mejorar aquello que no se mide de forma consistente, comparable y pública.

A lo largo de las últimas décadas, las evaluaciones a gran escala han dejado de ser instrumentos periféricos para ocupar un lugar central en el debate educativo. Han pasado a cumplir una función estratégica: hacer visible el nivel de aprendizaje de los estudiantes, orientar la toma de decisiones en políticas públicas y fortalecer el compromiso político con los resultados.

Al mismo tiempo, estas evaluaciones también conllevan límites y riesgos reales, que van desde costos elevados hasta distorsiones en los comportamientos y simplificaciones de la realidad educativa.

El objetivo de esta publicación es, precisamente, explorar esa tensión: reconocer los límites sin perder de vista el papel estructurante de las evaluaciones.

Este e-book busca responder a una pregunta central: ¿cómo utilizan —o dejan de utilizar— los distintos países las evaluaciones a gran escala para mejorar sus sistemas educativos?

PARTE I – EL PAPEL DE LAS EVALUACIONES

**Capítulo 1****Por qué importa medir el aprendizaje**

Para que los sistemas educativos, especialmente en América Latina, logren avanzar de forma consistente, es fundamental disponer de mecanismos que hagan visible lo que los estudiantes están aprendiendo, permitan seguir la evolución a lo largo del tiempo y creen condiciones para el ajuste de políticas públicas.

Las evaluaciones a escala no son políticas educativas en sí mismas, ni instrumentos pedagógicos de aula. Cumplen otra función: ayudan a estructurar el debate público, orientan decisiones a nivel sistémico y fortalecen el compromiso político con la mejora del aprendizaje.

A pesar de muchas experiencias que demuestran la potencia y la importancia de las evaluaciones de aprendizaje a escala, también existen limitaciones relevantes asociadas a este tipo de instrumento. Este artículo trata sobre estos límites, y sobre por qué, incluso reconociéndolos, las evaluaciones a escala continúan siendo fundamentales.

Antes, sin embargo, es importante delimitar el alcance de la discusión.



Existen usos inadecuados de las evaluaciones a gran escala. Hay actores que se apropian de estos instrumentos sin compromiso con la mejora de la educación, utilizando sus resultados para fines que poco tienen que ver con el aprendizaje. De la misma forma, hay rechazos a las evaluaciones que no se deben a limitaciones técnicas o pedagógicas, sino a la incomodidad con la idea de seguimiento, comparación y rendición de cuentas.

Estas posturas no parten de una mirada orientada a la mejora de los sistemas educativos. Por ello, no son el foco de este análisis. El debate que sigue parte del supuesto de que el objetivo es, de hecho, mejorar el aprendizaje, y hacerlo con equidad.

El costo de las evaluaciones a gran escala

Una crítica recurrente se refiere al costo. Las evaluaciones de aprendizaje a escala son caras, especialmente cuando están bien diseñadas e implementadas como deben serlo, y más aún cuando adoptan diseños censales. Este es un punto que merece ser reconocido sin rodeos.

Al mismo tiempo, la experiencia acumulada en diversos países muestra que estas evaluaciones pueden funcionar como un importante motor para la mejora de los sistemas educativos, las políticas públicas y, como consecuencia, los resultados de aprendizaje.

Bajo esta perspectiva, incluso con costos elevados, la inversión tiende a justificarse.

Además, existen caminos para mitigar estos costos. Uno de ellos se refiere al diseño de la evaluación. Cuando el objetivo central es el seguimiento y la generación de compromiso político, las evaluaciones muestrales pueden cumplir bien ese papel. Si un país o una entidad no dispone de recursos para una evaluación censal, es preferible realizar una evaluación muestral a simplemente no evaluar..

En países federativos, también existen sistemas institucionales alternativos. Las evaluaciones no necesitan ser exclusivamente nacionales. Pueden existir evaluaciones subnacionales, siempre que estén parametrizadas de modo a garantizar la comparabilidad. El Indicador Criança Alfabetizada en Brasil es un ejemplo de este tipo de esquemas.

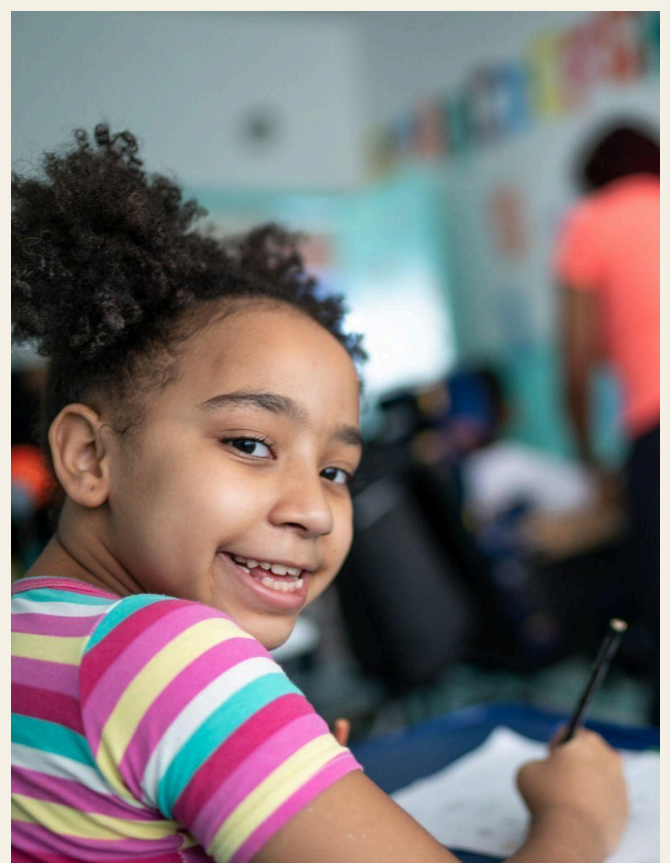
Otro camino relevante es el fortalecimiento de entidades técnicas especializadas en evaluación a gran escala. La actuación de socios técnicos contratados por el poder público puede generar ganancias de escala, eficiencia y reducción de costos, especialmente cuando estos actores operan en múltiples redes o territorios.

Por último, el uso de tecnología abre nuevas posibilidades. Las evaluaciones de fluidez con apoyo de IA, ampliamente utilizadas en alfabetización, son un ejemplo ya consolidado. Otras metodologías más recientes, como la evaluación ROAR (Rapid Online Reading Assessment), apuntan a caminos en los que la medición del aprendizaje puede volverse significativamente más barata.

La utilidad para el profesor y para el aula

Otra crítica frecuente es que las evaluaciones a gran escala poco ayudan al profesor en el trabajo cotidiano en el aula. En general, los resultados tardan en divulgarse, se presentan de forma agregada y no permiten diagnósticos individualizados de los estudiantes.

Esta crítica es, en parte, verdadera. Las evaluaciones a escala, la mayoría de las veces, no sustituyen a las evaluaciones diagnósticas e individualizadas, que son fundamentales para el trabajo pedagógico directo con el alumno. Idealmente, los sistemas educativos deberían contar con ambos tipos de instrumento.





Dependiendo del diseño, sin embargo, las evaluaciones a escala pueden sí generar información útil para escuelas y profesores, tema que se profundizará en el próximo artículo. Aun así, el punto central aquí es otro: las evaluaciones no necesitan cumplir todas las funciones al mismo tiempo.

El hecho de que una evaluación a escala no sirva directamente como instrumento diagnóstico no es un argumento en contra de su realización. Su papel principal, en este debate, es el seguimiento sistémico, el fortalecimiento del compromiso político y la orientación de políticas públicas. Incluso sin uso directo en el aula, puede hacer toda la diferencia para el funcionamiento del sistema en su conjunto.

El riesgo de estigmatización

La estigmatización es un riesgo real asociado a las evaluaciones a gran escala, especialmente cuando sus resultados se desagregan a niveles subnacionales, y cuando es posible, por escuela. Estos datos pueden ser usados para etiquetar a estudiantes, escuelas, profesores, gestores o redes, generando efectos negativos que poco contribuyen a la mejora de la educación.

Este riesgo, sin embargo, no está en la evaluación en sí, sino en el mal uso de sus resultados. Las utilidades con sesgo punitivo o exclusivamente negativo tienden a producir más efectos adversos que ganancias. La idea de que exponer resultados malos, por sí solo, llevará a la mejora de la educación parte de una visión simplista y poco realista sobre el funcionamiento de los sistemas educativos.

Las evaluaciones de aprendizaje no deberían ser utilizadas para estigmatizar. Este riesgo puede —y debe— ser enfrentado en el diseño y en la implementación de la evaluación.

Portanto, a possibilidade de estigmatização não é um argumento para abandonar avaliações em escala, mas um alerta para fazê-las melhor.



La simplificación de la realidad educativa

Las evaluaciones a gran escala son, por definición, una simplificación de la realidad. No capturan el desarrollo integral de niños y jóvenes. En general, se concentran en lengua y matemáticas, a veces en ciencias, dejando fuera diversas dimensiones centrales de la formación humana.

Esta limitación necesita ser reconocida. El problema surge cuando los resultados de la evaluación pasan a ser tratados como objetivo final, y no como una medición aproximada de la realidad. Mejorar el indicador no es sinónimo de mejorar la educación.

Por otro lado, como todo indicador, las evaluaciones a escala son útiles incluso siendo imperfectas. Hay evidencias consistentes de correlación entre los resultados de estas evaluaciones y dimensiones más amplias del desarrollo de los estudiantes. Además, miden conocimientos básicos que son condiciones necesarias, aunque no suficientes, para el desarrollo integral.

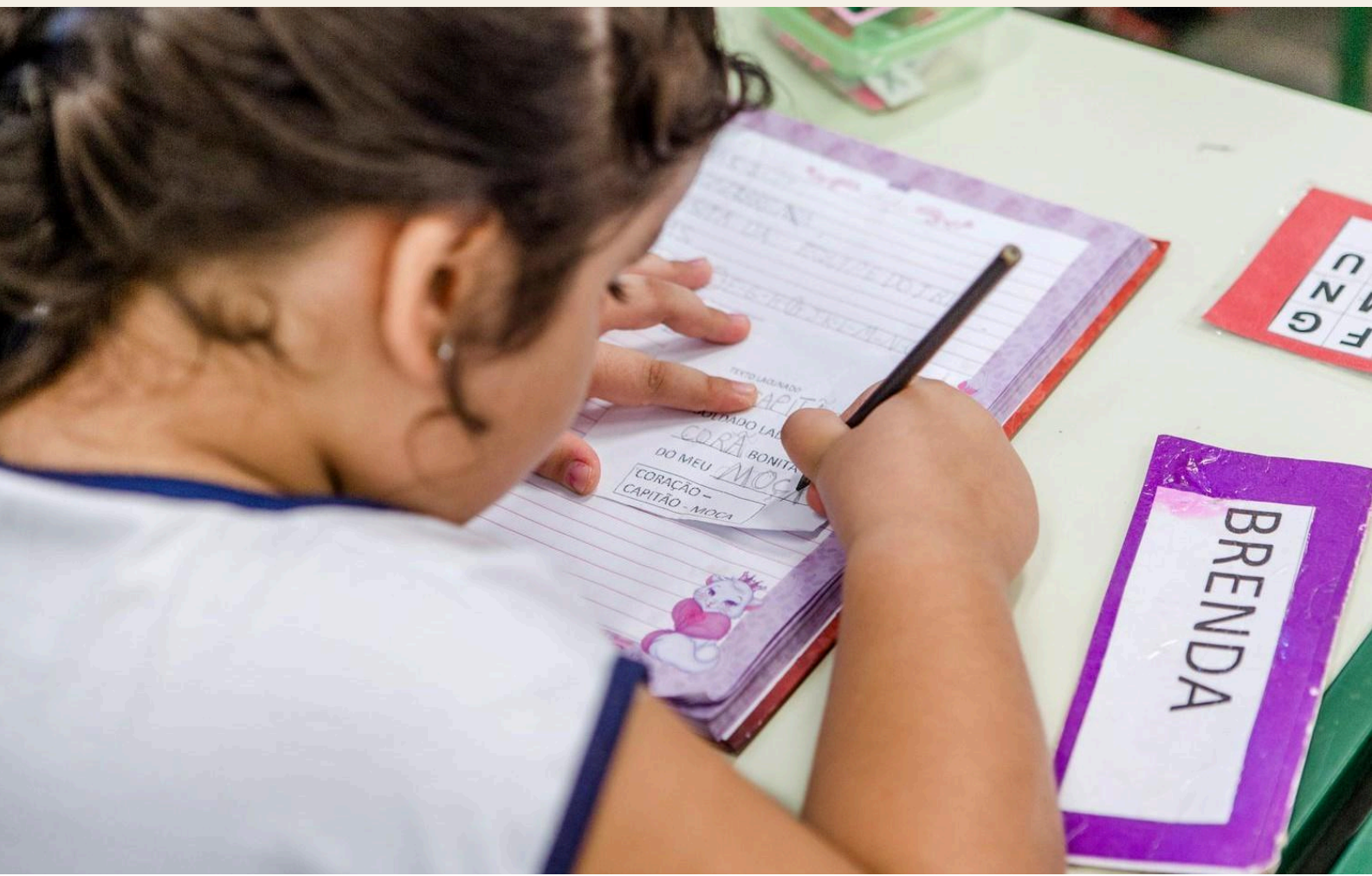
En el contexto de América Latina, donde los déficits en estos conocimientos aún son elevados, tiene mucho sentido medirlos. El indicador debe ser entendido como una lectura importante, pero simplificada de la realidad, y no como el estado ideal que se desea alcanzar.

Enseñanza enfocada en la prueba y distorsiones de comportamiento

Este punto está directamente relacionado con el anterior. Cuando los sistemas educativos pasan a entender que el objetivo del proceso educativo es ir bien en la evaluación, surgen distorsiones. La evaluación deja de ser instrumento y pasa a ser fin.

Estas distorsiones no siempre se deben a decisiones explícitas de “entrenar para la prueba”. Muchas veces son resultado de incentivos mal diseñados dentro del ecosistema educativo. El riesgo no está en utilizar evaluaciones a escala, sino en confundir su papel.

Las evaluaciones a escala miden un mínimo necesario, con correlación con el desarrollo integral, pero no agotan los objetivos de la educación. Es fundamental que el diseño y la implementación de estas evaluaciones incorporen mecanismos que reduzcan este tipo de efecto, tema que se tratará más adelante en el próximo artículo.



Ideologización de la evaluación de aprendizaje

Por último, existe la crítica de la ideologización. Las evaluaciones a gran escala pueden suscitar debates filosóficos legítimos sobre mérito, justicia, equidad y visiones de mundo. Estas discusiones son importantes y deben tener lugar.

Las evaluaciones de aprendizaje a escala tienen límites y riesgos reales. No capturan toda la complejidad de la educación, pueden generar distorsiones cuando se utilizan mal y plantean debates legítimos sobre costo, estigmatización y sentido pedagógico. Reconocer estas limitaciones es esencial para calificar el debate, y no para evitarlo.

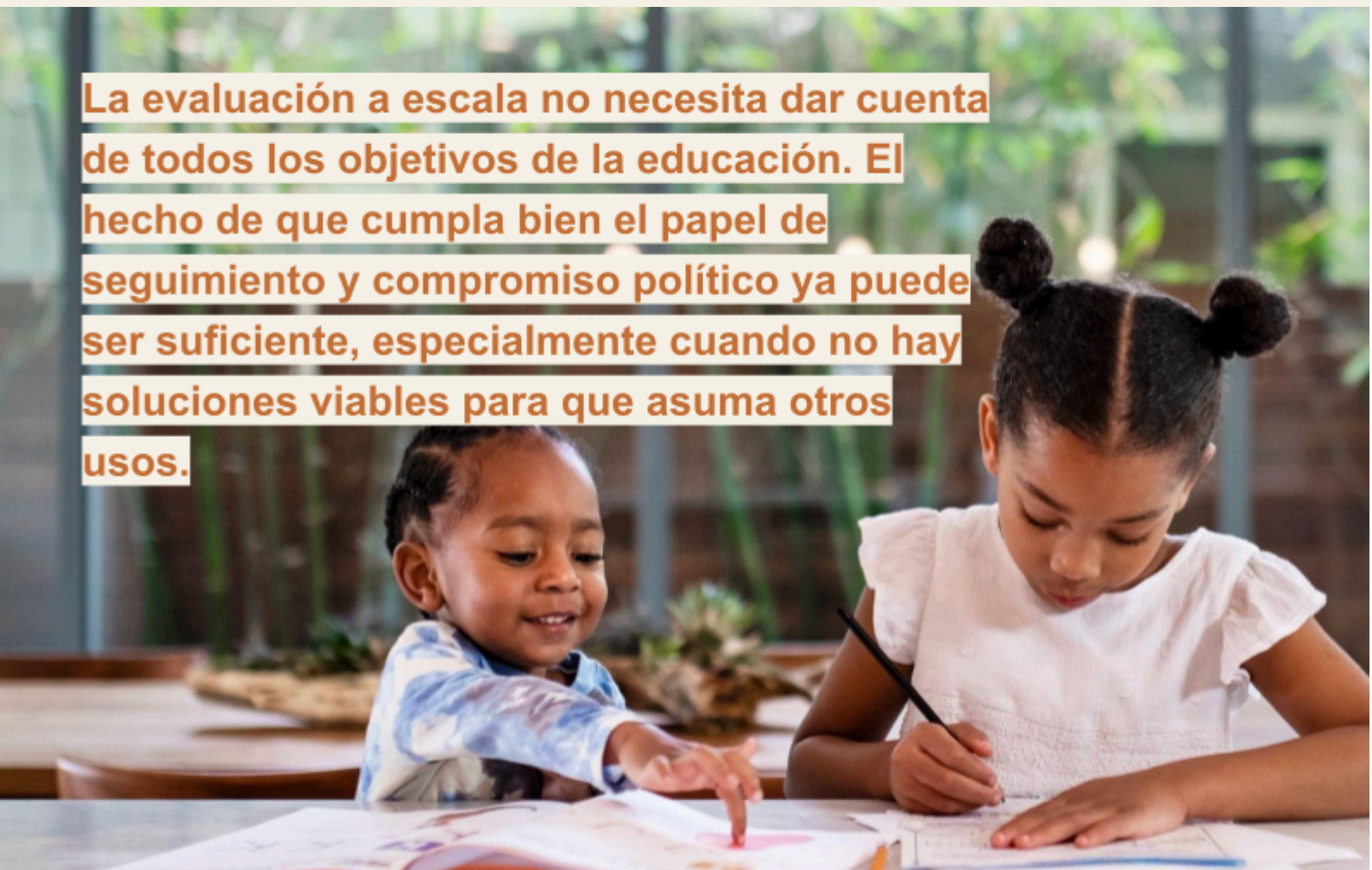


En esta publicación, discute estos límites de forma directa y por qué, aun así, las evaluaciones a gran escala siguen siendo fundamentales para monitorear sistemas educativos, orientar políticas públicas y fortalecer el compromiso político con la mejora del aprendizaje, especialmente en contextos marcados por desigualdades como el de América Latina.

Sin embargo, cuando el foco se desplaza al objetivo de mejorar el aprendizaje con equidad, la evaluación se muestra como una herramienta potente. Es difícil sostener, en términos sustantivos, el rechazo a objetivos como alfabetizar a todos los niños y niñas en la edad correcta o garantizar que los jóvenes concluyan la educación secundaria con un nivel básico de aprendizaje.

La experiencia internacional muestra que las evaluaciones a gran escala son adoptadas en países con orientaciones ideológicas muy distintas. Chile y Brasil, por ejemplo, bajo gobiernos de diferentes espectros políticos, mantuvieron y fortalecieron sus sistemas de evaluación. Esto sugiere que este instrumento puede —y debería— superar disputas partidistas.

La evaluación a escala no necesita dar cuenta de todos los objetivos de la educación. El hecho de que cumpla bien el papel de seguimiento y compromiso político ya puede ser suficiente, especialmente cuando no hay soluciones viables para que asuma otros usos.



Consideraciones finales

Reconocer que las evaluaciones de aprendizaje a escala tienen limitaciones es fundamental. Estos límites existen, son reales y deben tomarse en serio. Al mismo tiempo, muchos de estos riesgos pueden ser enfrentados mediante elecciones cuidadosas en el diseño y en la implementación de las evaluaciones.

Incluso con todas estas limitaciones, el poder de las evaluaciones a gran escala en términos de seguimiento, ajuste de políticas públicas, fortalecimiento del compromiso político e involucramiento de diferentes actores de la sociedad es lo suficientemente grande como para justificar su realización.

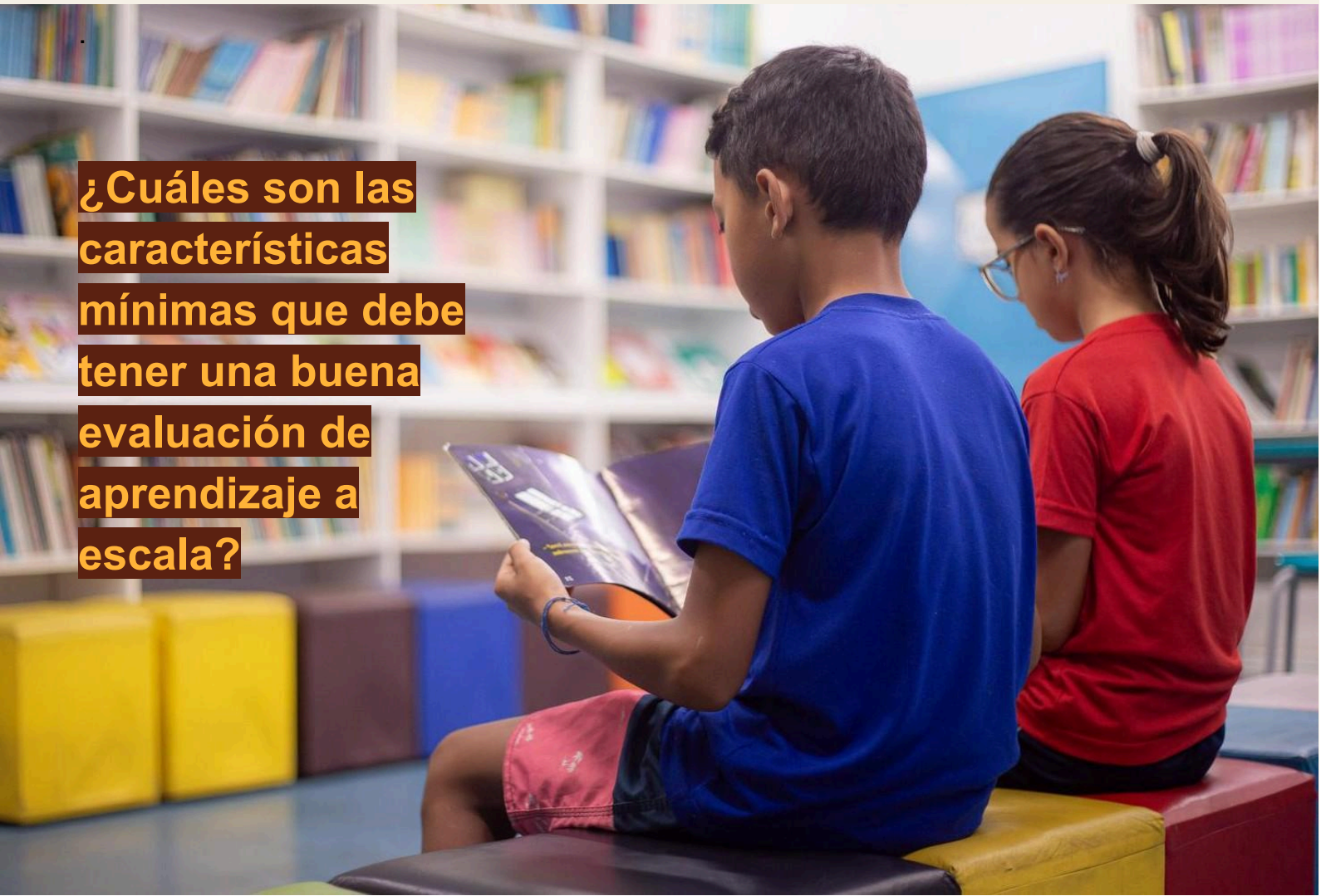
En el próximo capítulo, la discusión avanza a otro nivel: cuáles son las características de las evaluaciones de aprendizaje a escala que ayudan a mitigar estos riesgos y aumentan las posibilidades de que se utilicen efectivamente para mejorar el aprendizaje con equidad.

PARTE I – EL PAPEL DE LAS EVALUACIONES

Capítulo 2**¿Sirve cualquier evaluación de aprendizaje a escala?**

Una vez que consideramos la evaluación de aprendizaje a escala fundamental para generar y sostener el compromiso político —además de permitir la revisión y el ajuste de políticas públicas—, y aun reconociendo sus limitaciones, tema debatido en los dos últimos artículos, me gustaría centrarme, en este texto, en la discusión sobre qué características deben tener estas evaluaciones para cumplir adecuadamente su papel.

Vale recordar que las evaluaciones a escala pueden cumplir diferentes funciones. Aquí, sin embargo, me refiero específicamente a lo que considero su papel principal: apoyar la toma de decisiones en políticas públicas y fortalecer el compromiso político con la mejora del aprendizaje.

A photograph of two children sitting on a colorful cushioned bench in a library. The boy in the foreground is wearing a blue t-shirt and pink shorts, and is reading a book. The girl behind him is wearing a red t-shirt and glasses. The background shows bookshelves filled with books.

¿Cuáles son las características mínimas que debe tener una buena evaluación de aprendizaje a escala?

Objetivo

La evaluación de aprendizaje a escala debe existir, ante todo, con el objetivo de mejorar algún resultado educativo, en general mediante la implementación o la revisión de políticas públicas. Cuando no existe la intención real de actuar sobre los resultados, o, al menos, de llamar la atención sobre la necesidad de hacerlo, la evaluación se vuelve inútil y, a veces, contraproducente.

Curiosamente, el razonamiento inverso también suele ser verdadero: en ausencia de evaluaciones a escala con las características discutidas aquí, se vuelve difícil creer que el ente público esté genuinamente comprometido en mejorar los resultados educativos en cuestión.



Calidad técnica

Si deseamos medir, por ejemplo, el nivel de alfabetización en la edad correcta en un país, estado o municipio, es fundamental que el instrumento de evaluación sea técnicamente adecuado, es decir, que efectivamente mida lo que se propone medir.

Aunque esto parezca obvio, alcanzar este objetivo exige un esfuerzo significativo: definición clara de criterios, construcción de progresiones y elaboración de preguntas (u otros instrumentos) que sean, de hecho, capaces de evaluar el nivel de aprendizaje pretendido.

Comparabilidad en el espacio y en el tiempo

Si el objetivo es generar compromiso político y, como consecuencia, ajustes en políticas públicas, es fundamental que las evaluaciones sean comparables a lo largo del tiempo. La evaluación pierde gran parte de su valor —o puede incluso volverse sesgada, a veces con objetivos políticos— cuando no permite una lectura consistente de avances o retrocesos a lo largo de los años.



También es altamente deseable que haya comparabilidad en el espacio, entre estados, municipios o regiones. No se trata de crear clasificaciones, sino de comprender dónde las políticas han funcionado mejor o peor, permitiendo identificar lo que funciona, lo que no funciona y dónde hubo —o no— compromiso político.

No entraré aquí en detalles técnicos, pero cabe destacar la importancia tanto de la construcción de los ítems (por ejemplo, con el uso de la metodología TRI) como del diseño muestral y de otros aspectos metodológicos. Incluso las evaluaciones censales nunca alcanzan el 100% de participación, y garantizar que la población evaluada no esté distorsionada es un punto central.

Periodicidad

Para los objetivos mencionados, la repetición de la evaluación con una periodicidad adecuada es fundamental. Entiendo que la periodicidad máxima debería ser de dos años.

Reconozco que existe un debate según el cual los cambios en los resultados educativos tardan mucho tiempo y, por lo tanto, las evaluaciones en intervalos cortos serían poco informativas. Sin embargo, he observado la importancia de la periodicidad tanto para sostener el compromiso político como para captar avances más rápidos de lo esperado, especialmente en los años iniciales.

Un ejemplo es el Indicador Criança Alfabetizada (ICA), implementado en Brasil con periodicidad anual, que viene demostrando una tendencia de mejora. Si se aplicara cada cinco años, estos avances no se percibirían, y tampoco ocurrirían, con la misma intensidad, los ajustes en las políticas de alfabetización en régimen de colaboración o el refuerzo positivo a los gestores que están avanzando.



Confiabilidad y transparencia

Si el objetivo es alterar políticas públicas para alcanzar mejores resultados de aprendizaje —y si la implementación final de estos cambios ocurre en los sistemas locales y en las escuelas—, es fundamental que haya confianza en los resultados.

Esto implica asegurar que los datos sean reales, que no haya objetivos espurios y que no existan fallas metodológicas ocultas. Confiabilidad y transparencia son condiciones necesarias para que los resultados sean apropiados por los diferentes actores involucrados.

Publicación de los resultados

Este es un punto básico. Realizar evaluaciones de aprendizaje a escala y restringir el acceso a la información a un pequeño grupo de técnicos reduce significativamente su relevancia para los objetivos aquí discutidos.

Incluso considerando los riesgos de estigmatización, clasificaciones y uso inadecuado de los datos —que deben mitigarse tanto como sea posible—, entiendo que, para los fines descritos en este artículo, la publicación de los resultados debe ser lo más amplia posible.

También es importante que los resultados sean comprensibles para un público más amplio, y no solo para especialistas. Retomando el ejemplo del ICA: divulgar que el promedio nacional fue de 560 puntos (número ilustrativo) dice poco para la mayoría de las personas. Informar que, en 2023, el 59,2% de los niños estaban alfabetizados en la edad correcta, comunica de forma mucho más clara el significado del resultado.



Uso adecuado

Por último, como cualquier instrumento, las evaluaciones de aprendizaje a escala pueden ser mal utilizadas. Pueden servir solo para castigo, ser instrumentalizadas en disputas políticas o contribuir a la estigmatización de determinadas regiones.

Es necesario trabajar para mitigar estos riesgos. Aun así, como argumenté en el capítulo anterior, incluso considerando estas posibilidades de mal uso, entiendo que las evaluaciones de aprendizaje a escala son fundamentales para el avance de la calidad de la educación en los países latinoamericanos.



Nota final sobre terminología

En los textos sobre evaluación de aprendizaje a escala, utilizo el término ***evaluación de aprendizaje*** para referirme a lo que parte de la literatura especializada denomina evaluación de desempeño. Tengo claridad de esta distinción conceptual. La opción por el término aprendizaje fue deliberada, por considerarlo más accesible y menos confuso para el público más amplio con el que dialogo aquí, especialmente en el debate sobre políticas públicas y sistemas educativos.

Esto no significa confundir desempeño con aprendizaje, ni atribuir a las evaluaciones externas un papel que no tienen. A lo largo de los textos, busco dejar claro que estas evaluaciones observan el desempeño en tareas estructuradas y producen evidencias que permiten inferencias sobre el aprendizaje, sin jamás sustituir el seguimiento pedagógico realizado en el día a día del aula.

La distinción conceptual es importante y exactamente por eso vale explicitarla.

PARTE II – SISTEMAS NACIONAIS EN PERSPECTIVA

Capítulo 3**Brasil: solidez y complementariedad**

Como dije, las evaluaciones de aprendizaje a escala son instrumentos centrales para el seguimiento de los resultados educativos, para la generación de compromiso político y para el ajuste de políticas públicas, aun cuando es fundamental reconocer sus límites y riesgos

A partir de ahora, doy un paso adicional. Me uno a Karina Stocovaz, Gerente Sr. del Instituto Natura para los países de habla hispana, para realizar un análisis comparativo de los países donde actúa el Instituto Natura, con foco en las evaluaciones de aprendizaje a escala.

El análisis se basa en un conjunto de criterios para la evaluación de instrumentos, incluyendo calidad técnica, comparabilidad, transparencia, institucionalización, uso coherente de tecnología e incidencia en políticas públicas. Al final de los capítulos, presentaremos los análisis en formato de tabla, de una manera más objetiva, lo que permitirá algún tipo de comparabilidad entre los diferentes países..



Comenzamos con un panorama de las evaluaciones nacionales de aprendizaje a gran escala existentes hoy en Brasil, observando sus principales características, usos y grados de consolidación institucional.

Actualmente, Brasil cuenta con dos evaluaciones nacionales recurrentes y a gran escala:

- el SAEB, enfocado en las etapas finales de la Educación Primaria y la Educación Secundaria;
- el SAEB Alfabetização, centrado en los años iniciales, especialmente en el 2º año de la Educación Primaria.

Aunque forman parte de un mismo sistema, estas evaluaciones cumplen roles distintos y presentan diferentes niveles de madurez institucional.

SAEB Alfabetização: foco reciente, alto uso en política pública

El SAEB Alfabetização es una evaluación relativamente reciente, iniciada en 2021, enfocada en la medición de la alfabetización inicial, con foco en lectura y escritura al final del 2º año de la Educación Primaria.

Desde el punto de vista técnico, existe cierto consenso sobre su importancia y su validez como instrumento hasta ahora; no obstante, hay un debate muy fuerte entre los especialistas sobre la necesidad de revisar el modelo de la evaluación para que se alinee a innovaciones metodológicas de evaluaciones internacionales. La transparencia es alta, con amplia disponibilidad de microdatos, informes y documentación metodológica.

La evaluación tiene periodicidad anual, característica que la hace especialmente adecuada para el monitoreo continuo de las políticas de alfabetización. Sus resultados son ampliamente comparables entre estados y municipios y se han divulgado de forma clara, con indicadores fácilmente comprensibles para públicos no especializados, como el porcentaje de niños alfabetizados en la edad correcta.

Por otro lado, por ser un instrumento reciente, el SAEB Alfabetização aún posee una historia corta, lo que limita los análisis a largo plazo. Su grado de institucionalización puede considerarse intermedio: aunque está plenamente integrado a la política nacional de alfabetización, todavía se encuentra en proceso de consolidación como política de Estado.

Un punto a destacar es el uso efectivo de sus resultados en políticas públicas, especialmente en el régimen de colaboración entre la Unión, los estados y los municipios, lo que refuerza su papel como instrumento de monitoreo y coordinación federativa.

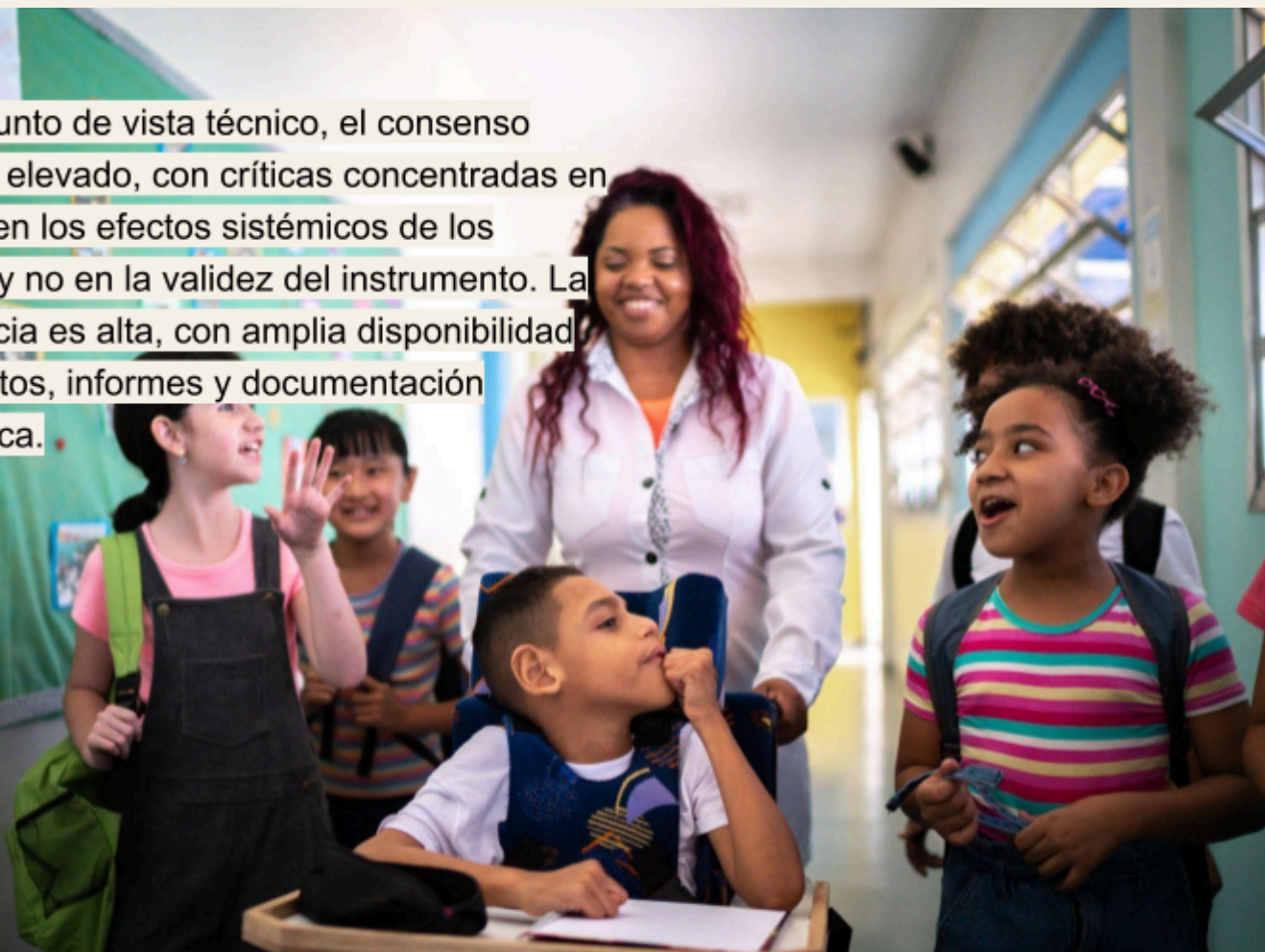


SAEB: trayectoria larga y papel estructurante en el sistema educativo

El SAEB, a su vez, es una de las evaluaciones a gran escala más consolidadas de la región. Aplicada desde 1990, abarca el 5° y el 9° año de la Educación Primaria y el 3° año de la Educación Secundaria, con foco principal en Lengua Portuguesa y Matemática.

Se trata de una evaluación con alto grado de institucionalización, continuidad a lo largo de diferentes gobiernos y por una extensa serie histórica, lo que permite lecturas consistentes de tendencias a lo largo del tiempo. La comparabilidad territorial es elevada, con amplia desagregación de los resultados por estados, municipios y redes.

Desde el punto de vista técnico, el consenso también es elevado, con críticas concentradas en los usos y en los efectos sistémicos de los resultados y no en la validez del instrumento. La transparencia es alta, con amplia disponibilidad de microdatos, informes y documentación metodológica.



Los resultados del SAEB son ampliamente divulgados y utilizados en políticas públicas, incluso en instrumentos como el IDEB, que combina escalas de rendimiento en una métrica más simple con tasas de aprobación.

Por otro lado, a diferencia del SAEB Alfabetização, los resultados del SAEB, en su forma original, exigen una mayor mediación técnica para ser comprendidos por el público general, lo que limita su lectura directa fuera de círculos especializados.

Dos evaluaciones, roles complementarios

El panorama brasileño revela un sistema de evaluación a gran escala relativamente sólido, con instrumentos que se complementan:

- El SAEB, más consolidado, cumple un papel estructurante de monitoreo sistémico a lo largo del tiempo.
- El SAEB Alfabetização, más reciente, ha demostrado ser particularmente potente para inducir la coordinación, el foco y ajustes rápidos en políticas de alfabetización.

Ambas presentan un alto grado de uso en políticas públicas, amplia publicidad y fuerte comparabilidad territorial. Las diferencias se concentran sobre todo en el tiempo de maduración institucional, la extensión de la serie histórica y la forma en que los resultados dialogan con el público no especializado.

Este conjunto de características ayuda a explicar por qué las evaluaciones a gran escala ocupan hoy un lugar central en el debate educativo brasileño, y también señala los desafíos que persisten para su perfeccionamiento.



PARTE II – SISTEMAS NACIONAIS EN PERSPECTIVA

Capítulo 4**Argentina: tradición y desafíos de consolidación**

Argentina cuenta actualmente con dos evaluaciones nacionales a gran escala en la educación básica:

- Aprender, aplicada principalmente en el 6° año de la educación primaria.
- Aprender Alfabetización, enfocada en el 3° año e insertada en el marco del Plan Nacional de Alfabetización.

Aunque forman parte del mismo sistema, se encuentran en distintas etapas de madurez institucional y de consolidación como instrumentos estructurantes de política pública.

Aprender: tradición evaluativa y debates persistentes

La evaluación Aprender posee una trayectoria relativamente consolidada desde 2016, con posibilidad de comparación con la edición anterior (Operativo Nacional de Evaluación, de 2013). De hecho, Argentina acumula experiencia evaluativa desde la década de 1990, lo que confiere densidad técnica al sistema.

Aplicada de forma censal en el 6° año de la educación primaria —con alternancia para la educación secundaria—, la evaluación permite una amplia desagregación por jurisdicción, sector de gestión y grupos poblacionales. Esta característica es particularmente relevante en un país federal como Argentina, donde las provincias tienen un papel central en la gestión educativa.



La metodología ha buscado mantener estabilidad a lo largo del tiempo, lo que posibilita la lectura de tendencias. Aun así, el instrumento es objeto de debates recurrentes. Parte de las discusiones gira en torno a la capacidad de la evaluación para captar procesos más amplios de aprendizaje; otra parte se refiere a la demanda por divulgación de resultados individualizados por alumno o escuela, algo que no corresponde al diseño metodológico de la prueba, construida para lectura sistémica y no para medición individual.



La transparencia de los datos se logra mediante la publicación de informes nacionales y jurisdiccionales, microdatos (con preservación de la confidencialidad), notas metodológicas y plataformas dinámicas de consulta. Los resultados son ampliamente divulgados y frecuentemente tienen repercusión en el debate educativo y en los medios.

Al mismo tiempo, la estructura institucional responsable de la evaluación está vinculada directamente a la Secretaría de Educación. Aunque existe una tradición técnica consolidada, no hay una plena protección institucional frente a cambios de gobierno, lo que podría generar inestabilidades a lo largo del tiempo.

En lo que se refiere al uso en políticas públicas, los resultados se utilizan para identificar escuelas que demandan apoyo y para fundamentar iniciativas de refuerzo de aprendizajes. Sin embargo, este uso no se presenta de forma plenamente sistemática o estructurante en el ciclo de formulación de políticas nacionales.

Aprender Alfabetización: una línea de base reciente

Aprender Alfabetización se aplicó por primera vez en 2024, con foco en el 3º año de la educación primaria e insertada en el Plan Nacional de Alfabetización.

Se trata de una evaluación nueva, todavía en fase inicial de consolidación. Su aplicación tuvo como objetivo establecer una línea de base para la política nacional de alfabetización, con previsión de una nueva medición en 2026 para el seguimiento de los avances.

Al haber introducido cambios metodológicos relevantes en relación con la evaluación anterior aplicada en el 3º año, no hay comparabilidad directa con ediciones pasadas. Esto limita, por ahora, la lectura de tendencias en el tiempo.

La evaluación fue muestral, con representatividad por jurisdicción, sector de gestión, nivel socioeconómico y sexo, lo que permite análisis relevantes, aunque con un menor grado de detalle territorial que una aplicación censal.

Para garantizar la transparencia, se divulgaron informes nacionales y jurisdiccionales, recomendaciones pedagógicas, cuestionarios de contexto y bases de datos públicas.

Uno de los principales debates se centró en la definición del punto de corte que caracterizaría a un niño como alfabetizado. Se divulgaron múltiples niveles de desempeño, pero la ausencia de un marco público claro que sintetice la información en una métrica simple dificulta la comunicación de los resultados para públicos no especializados.



Por otro lado, el uso de la evaluación como línea de base para la política nacional de alfabetización es un elemento relevante. La propia previsión de una nueva aplicación con el objetivo explícito de medir avances asociados a la política indica una intención de vincular evaluación y acción pública de manera más directa.

Aun así, se trata de un instrumento con un historial muy corto y con una continuidad institucional todavía en proceso de definición.

Un sistema con bases sólidas, pero desafíos de consolidación

El panorama argentino revela un sistema evaluativo con tradición técnica, alta transparencia y amplia divulgación pública. La evaluación Aprender presenta una mayor estabilidad histórica e institucional, mientras que Aprender Alfabetización cumple un papel estratégico reciente en el ámbito de la política de alfabetización.

Las diferencias entre ambas se concentran sobre todo en la madurez institucional, la extensión de la serie histórica y la forma en que los resultados se traducen para el debate público y se incorporan al ciclo de políticas.

En general, Argentina cuenta con instrumentos que permiten una lectura consistente del sistema educativo. El desafío reside en garantizar la continuidad de las evaluaciones, con comparabilidad en el tiempo —especialmente en el caso de la alfabetización— y avanzar en la integración más sistemática entre evaluación y toma de decisiones.

Como veremos en los próximos países de la serie, las evaluaciones a gran escala ocupan lugares distintos en los sistemas educativos latinoamericanos. Entender estas diferencias es fundamental para comprender no sólo los resultados, sino también la capacidad de cada país para transformar datos en política pública consistente.

PARTE II – SISTEMAS NACIONAIS EN PERSPECTIVA

**Capítulo 5****Perú: institucionalización con tensiones metodológicas**

Perú consolidó, a lo largo de las últimas décadas, un sistema de evaluación educativa que se convirtió en un pilar importante para el monitoreo de políticas públicas. Con una trayectoria que se remonta a 1996, el país consolidó la Evaluación Censal de Estudiantes (ECE) como principal instrumento nacional de monitoreo del aprendizaje, con aplicaciones regulares entre 2007 y 2019.

A partir de 2023, Perú pasó a organizar sus mediciones bajo un marco más amplio, la Evaluación Nacional de Logros de Aprendizaje (ENLA), que articula diferentes modalidades de evaluación —censales y muestrales— y amplía el alcance analítico del sistema.

En este nuevo esquema, la ENLA no sustituye necesariamente a la ECE, sino que funciona como un paraguas evaluativo, incorporando aplicaciones censales cuando son pertinentes y ampliando el análisis a factores de contexto —incluyendo dimensiones socio emocionales— que influyen en la vida escolar cotidiana.

Aunque el sistema presenta un alto grado de institucionalización y continuidad entre diferentes gobiernos, enfrenta desafíos estructurales y metodológicos que impactan la lectura precisa de los avances educativos a largo plazo.

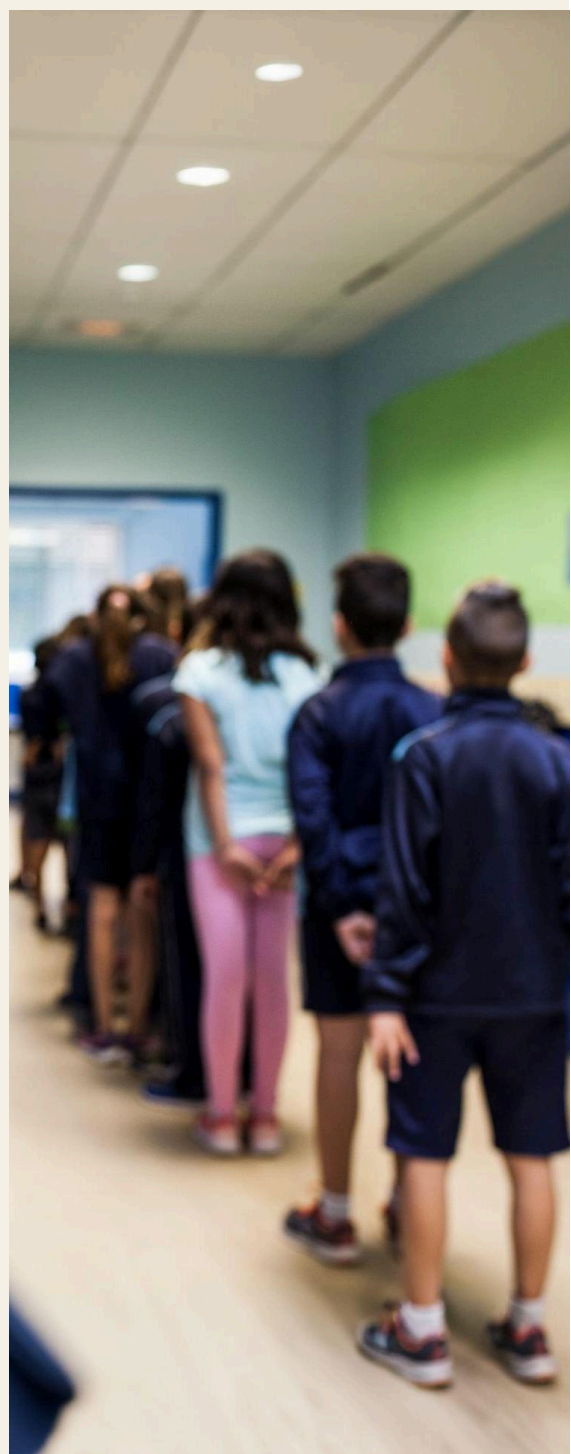


Esta reorganización aumenta la flexibilidad del sistema, pero también exige cautela en la lectura de tendencias a largo plazo, ya que los cambios de diseño y cobertura pueden afectar la comparabilidad directa con series anteriores.

Calidad técnica y los desafíos de la comparabilidad

La calidad técnica de los instrumentos peruanos es reconocida, presentando un alto grado de consenso nacional. La metodología es considerada estable y está alineada al Currículo Nacional de la Educación Básica. No obstante, especialistas señalan puntos de atención que pueden afectar la comparabilidad en el espacio y en el tiempo —especialmente en el contexto de la reciente reorganización del sistema evaluativo—, factor que consideramos fundamental.

- **Alteración en la definición de ruralidad:** En 2021, el Ministerio de Educación (Minedu) modificó los criterios de lo que se considera zona rural. Este cambio incluyó un número significativamente mayor de escuelas en esta categoría, lo que puede comprometer la comparabilidad directa de los resultados actuales con las ediciones anteriores.
- **Representatividad territorial limitada:** En los años en que la evaluación se realiza de forma muestral (EM), los datos no poseen representatividad para el nivel de UGEL (Unidades de Gestión Educativa Local). Esto crea una laguna para los gestores de primera línea, que solo obtienen datos específicos para sus jurisdicciones en los ciclos censales —modalidad históricamente asociada a la ECE y que sigue utilizándose dentro del marco ENLA.
- **Independencia institucional:** La UMC (Oficina de Medición de la Calidad), responsable de las mediciones, está vinculada directamente al Ministerio. A pesar de su tradición técnica, esta estructura no ofrece una protección completa contra posibles cambios de prioridades asociados a los ciclos políticos.





Inclusión y Diversidad: El desafío de la EIB1

Un diferencial relevante en el proceso evaluativo de Perú es el enfoque en la Educación Intercultural Bilingüe (EIB). Desde 2007, el país realiza evaluaciones de lectura en lenguas originarias para alumnos de 4° año, abarcando idiomas como Quechua (en diversas variantes), Aimara, Awajún y Shipibo-Konibo.

Esta medición es fundamental en un país con tanta diversidad cultural, pero impone desafíos logísticos y técnicos complejos. La evaluación busca garantizar que el sistema no ignore las particularidades lingüísticas de una parte significativa de la población, aunque la escala de estas aplicaciones y la producción de materiales específicos todavía enfrentan barreras de implementación a gran escala.

La trayectoria de los años evaluados y el uso de la tecnología

La serie histórica peruana revela una transición de foco a lo largo de los años. Entre 2007 y 2014, el sistema se concentró casi exclusivamente en el 2° año de la Educación Primaria. Fue solo en 2016 que el 4° año pasó a ser evaluado de forma censal para la enseñanza regular, convirtiéndose, desde entonces, en la serie de referencia para el monitoreo continuo. Considerando el momento en que varios países de América Latina están prestando más atención a la alfabetización inicial, incluido Perú, no tener evaluaciones con periodicidad y comparabilidad en el 2° año es, ciertamente, un problema.

En lo que respecta a la tecnología, Perú demuestra una etapa intermedia de transición. Durante los años de pandemia, el país implementó la EVA (Evaluación Virtual de Aprendizajes). Aunque fue una respuesta necesaria al cierre de las escuelas, el uso de plataformas digitales todavía se ve con cautela para garantizar que la falta de infraestructura tecnológica en áreas remotas no profundice las desigualdades ya existentes en los resultados.

Transparencia y la "Educación en Datos"

El sistema peruano dedica esfuerzos a la democratización del acceso a la información. Los datos se traducen en escalas de niveles de logro —como "Satisfactorio", "En proceso" y "En inicio"— que buscan hacer que el dato estadístico sea más inteligible para las familias. El Minedu utiliza recursos diversificados, como podcasts y seminarios web, para ampliar el alcance del debate. Además, existe un portal de consulta dinámica que permite el acceso a microdatos históricos, integrando resultados nacionales e internacionales.



La incidencia de los resultados en las políticas públicas

El nivel de incidencia de las evaluaciones en el diseño de políticas públicas se considera medio. En el pasado, los resultados fundamentaron decisiones como el Bono de Incentivo al Desempeño Escolar (BDE), que bonificaba a docentes y directores de escuelas con avances significativos. Los datos también orientaron el Programa Estratégico Logros de Aprendizaje (PELA), que dirigió apoyo pedagógico y recursos presupuestarios a las regiones con mayores lagunas de aprendizaje. Más recientemente, el país pasó a difundir "Kits de Evaluación Diagnóstica", permitiendo que los profesores utilicen los parámetros de la ENLA para monitorear el progreso de los alumnos directamente en el aula. Actualmente, la claridad sobre el uso efectivo de las evaluaciones en el diseño y seguimiento de las políticas educativas se ha disipado. Esta situación constituye un verdadero desafío y una oportunidad para el país.

PARTE II – SISTEMAS NACIONAIS EN PERSPECTIVA

Capítulo 6**Chile: madurez, crisis y reinención**

Si hay un país en la región que ejemplifica la maduración institucional de un sistema de evaluación —pasando por fases de aceptación técnica absoluta, crisis de legitimidad y, finalmente, una sofisticada modernización tecnológica—, ese país es Chile.



El sistema chileno de evaluación a gran escala se articula en torno a un instrumento fundamental: el SIMCE (Sistema de Medición de la Calidad de la Educación). A diferencia de otros vecinos que aún luchan por consolidar series históricas, Chile ostenta una tradición que se remonta a 1988. Sin embargo, el SIMCE no es solo un sobreviviente del tiempo; es un sobreviviente de las transformaciones políticas y sociales del país. Su trayectoria está marcada por un desplazamiento profundo: de un modelo de alta presión, que en los años 90 era visto con gran confianza por su precisión técnica, a un escenario contemporáneo de intenso debate sobre el "para qué" y el "cómo" evaluamos.

Esta evolución fue impulsada por algunos cuestionamientos. En el pasado, el uso excesivo de los resultados para fines de clasificación y sanción administrativa generó efectos colaterales severos, como por ejemplo el empobrecimiento del currículo escolar, ya que muchas instituciones, presionadas por los números, pasaron a "enseñar para el examen", lo que en la práctica llevó a la reducción de tiempo en algunas disciplinas para priorizar Matemática y Lenguaje. Aunque la rigidez del currículo actual limita esta práctica, la crítica sobre el sacrificio curricular es una parte importante de la historia del debate evaluativo de aquel periodo.

Para salvar la legitimidad del instrumento y evitar que se convirtiera en un mecanismo de segregación, Chile promovió una reforma esencial. Hoy, el sistema busca un mayor equilibrio. La evaluación no se restringe solo al SIMCE, integrándose a un sistema más amplio que incluye otras herramientas, reduce el número de etapas evaluadas e incorpora los llamados Indicadores de Desarrollo Personal y Social, señalando que la calidad educativa no se resume solo al rendimiento académico. También se puso a disposición de las escuelas una herramienta de evaluación diagnóstica digital que devuelve resultados inmediatos a los docentes (DIA, Diagnóstico Integral de Aprendizaje).



Uno de los puntos de mayor destaque en el caso chileno, y que sirve de referencia para toda América Latina, es su pionerismo tecnológico.


Chile opera a la vanguardia digital de la evaluación, con un nivel de sofisticación que permite reducir el ciclo de procesamiento de datos de 220 a solo 120 días. Esta agilidad no es solo burocrática; garantiza que el dato llegue a la escuela aún con frescura pedagógica. Así, por ejemplo, pocos días después de iniciado el año lectivo, ya se publicaron los resultados del SIMCE 2025. Esta entrega rápida ya ocurre desde hace cuatro años consecutivos.

El uso de algoritmos de inteligencia artificial para detectar anomalías y la creación de portales de resultados extremadamente intuitivos y amigables para padres y directores transforman el SIMCE en una herramienta de gestión viva. Además, el país dio grandes pasos en la inclusión, utilizando software de adaptación y lectores de pantalla para que estudiantes con discapacidad puedan ser evaluados con equidad.

En términos de incidencia en las políticas públicas, el SIMCE es el pilar del sistema de rendición de cuentas chileno. Si en Argentina el uso de los resultados aún busca una mayor sistematización en las decisiones políticas, en Chile es estructurante. No obstante, es necesaria cautela para no sobredimensionar su papel, que, aunque importante, forma parte de un conjunto mayor de evidencias y acciones políticas. Los datos alimentan directamente a la Agencia de Calidad de la Educación, un organismo que goza de significativa autonomía y que utiliza las evidencias para clasificar y apoyar a las escuelas.



La institucionalización del sistema chileno también merece destaque. La evaluación es un pilar de la Ley N° 20.529 y es ejecutada por la Agencia de Calidad de la Educación, un organismo que goza de autonomía y estabilidad técnica, protegiendo al sistema de variaciones políticas bruscas. Contrariamente a lo que observamos en algunos contextos donde la evaluación es vista como una herramienta punitiva, Chile ha buscado transformar el SIMCE en un mecanismo de apoyo. El sistema clasifica a las escuelas y dirige recursos y soporte técnico para aquellas en situación crítica, haciendo que la incidencia en políticas públicas sea extremadamente alta y estructurante.



Con más de 22 años de series históricas comparables (desde 2002), Chile logra comunicar sus resultados de forma muy clara para el público no especializado.

Las métricas de rendimiento se traducen en niveles intuitivos —Insuficiente, Elemental y Adecuado—, lo que permite un control social más efectivo por parte de la comunidad escolar. Al mirar a Chile, vemos un sistema veterano que supo reconocer sus límites y modernizarse, demostrando que la tecnología y la transparencia son los mejores antídotos para el desgaste de un sistema de evaluación a gran escala.

En conclusión, el caso chileno se erige como un referente de madurez institucional y resiliencia en América Latina. A pesar de su compleja trayectoria, el sistema de evaluación chileno demostró un compromiso con la transparencia, la institucionalización robusta y un notable pionerismo tecnológico, pilares que garantizan datos confiables, ciclos de procesamiento ágiles y una comunicación clara de resultados al público no especializado.

Estos cimientos transforman al SIMCE en una herramienta de gestión viva y de alto impacto en las políticas públicas. No obstante, los resultados recientes, como el estancamiento en Lectura y el alto porcentaje de estudiantes en el nivel 'Insuficiente' en áreas clave, nos recuerdan que la persistencia del desafío estructural es innegable. La gran lección de Chile es que, si bien la tecnología y la solidez institucional son antídotos contra el desgaste del sistema, el éxito final se medirá por su capacidad de traducir la evidencia en una mejora pedagógica efectiva que impulse la calidad y la equidad para todos los estudiantes.

PARTE II – SISTEMAS NACIONAIS EN PERSPECTIVA

Capítulo 7**Colombia: sistema estructurado, consistencia y el desafío de la última milla**

Colombia cuenta actualmente con un sistema consolidado de evaluación nacional de aprendizaje a gran escala, conocido como Saber, que mide los resultados educativos en diferentes etapas de la trayectoria escolar.

Saber: un sistema estructurado de evaluación del sistema educativo

Las evaluaciones Saber constituyen el principal instrumento nacional de medición de aprendizaje en Colombia. Se aplican en diferentes momentos de la trayectoria escolar y forman parte de un sistema integrado de evaluación educativa conducido por el ICFES (Instituto Colombiano para la Evaluación de la Educación).



Desde el punto de vista técnico, las evaluaciones Saber poseen un alto grado de reconocimiento entre especialistas, gestores e investigadores, con metodología consolidada y continuidad a lo largo del tiempo. Las críticas existentes se concentran más en el uso pedagógico de los resultados y en la periodicidad de la aplicación..

Otro aspecto relevante es la amplia comparabilidad territorial. Los resultados se divulgan con desagregaciones por departamentos, municipios y tipos de escuela, lo que permite análisis detallados de las desigualdades educativas dentro del país, una característica especialmente importante en sistemas educativos descentralizados.



Comparabilidad histórica y transparencia de los datos

Uno de los puntos fuertes del sistema colombiano es la continuidad de las mediciones a lo largo del tiempo, lo que permite la construcción de series históricas y la identificación de tendencias en el rendimiento educativo.

La estabilidad metodológica y la regularidad de las aplicaciones permiten que investigadores, gestores públicos y organizaciones de la sociedad civil acompañen la evolución de los resultados con relativa consistencia.

La transparencia de los datos también es elevada. El sistema pone a disposición informes técnicos, documentación metodológica y bases de datos que permiten análisis más profundos. La divulgación pública de los resultados suele ocurrir de forma amplia y regular, con repercusión en el debate educativo y en los medios.

Otro elemento relevante es la forma de comunicación de los resultados. Aunque las evaluaciones utilizan escalas técnicas de rendimiento, el sistema también traduce los resultados en niveles de desempeño que facilitan la interpretación para gestores y para el público más amplio.



Institucionalización y gobernanza del sistema de evaluación

El sistema Saber presenta un alto grado de institucionalización, con continuidad a lo largo de diferentes gobiernos y papel consolidado en la arquitectura de las políticas educativas colombianas.

La responsabilidad técnica por la evaluación es ejercida por el ICFES, una institución con tradición técnica en el país. Aunque está vinculada al gobierno, la organización ha construido a lo largo del tiempo una reputación de capacidad técnica y estabilidad institucional.

Esta combinación entre tradición técnica, continuidad de las mediciones y divulgación pública de los resultados contribuye a que el sistema de evaluación desempeñe un papel estructurante en la producción de información sobre la educación colombiana.

Uso de los resultados y relación con políticas públicas

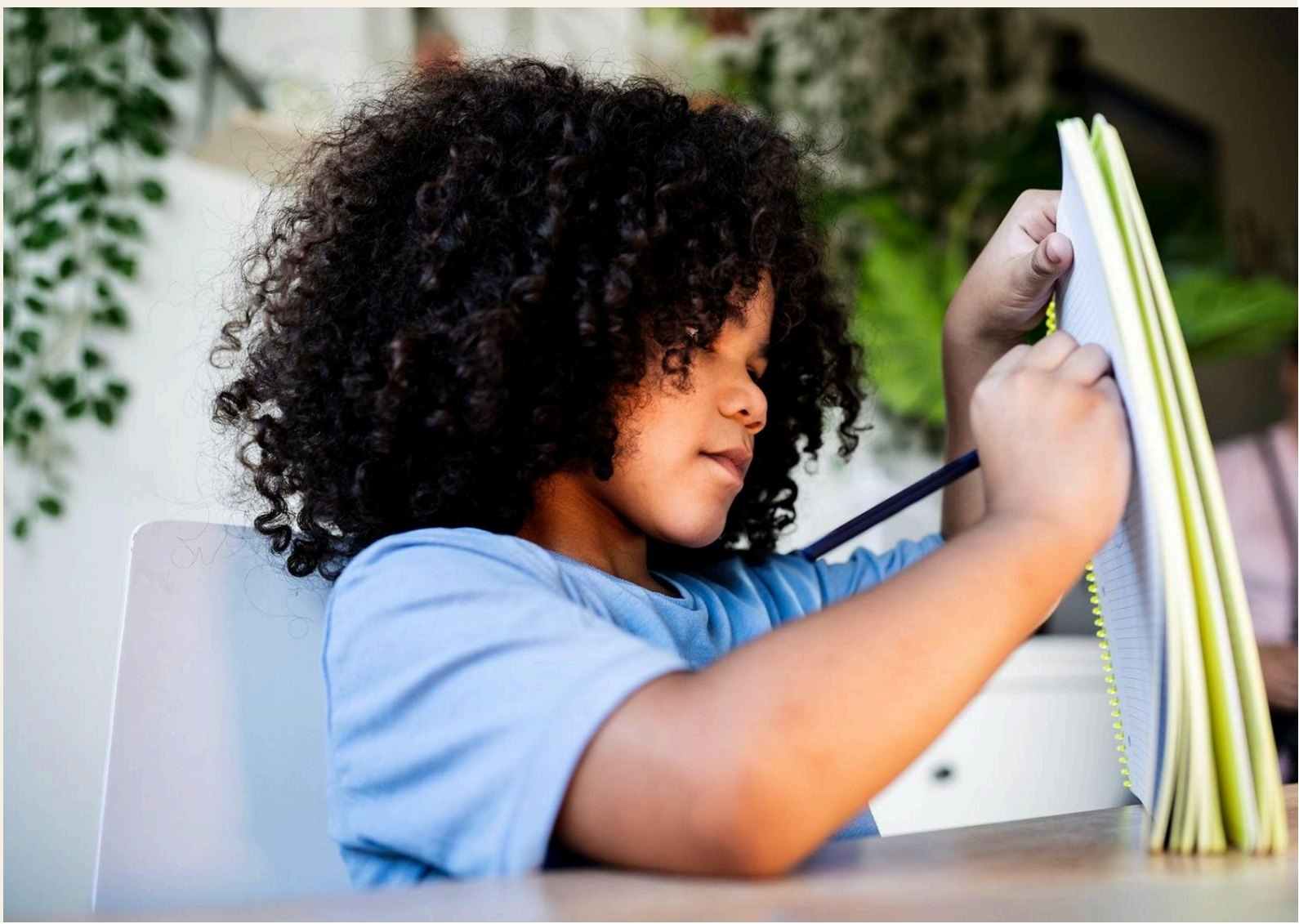
Los resultados de las evaluaciones Saber se utilizan para diversas finalidades dentro del sistema educativo colombiano.

Entre los usos más relevantes están:

- el monitoreo del desempeño educativo en diferentes niveles del sistema;
- la identificación de desigualdades territoriales de aprendizaje;
- el apoyo al diseño y ajuste de políticas educativas.

Este uso relativamente estructurado de la información refuerza el papel de la evaluación como instrumento de gestión educativa y de rendición de cuentas pública.

Al mismo tiempo, como ocurre en muchos países, persiste el desafío de profundizar la traducción de los resultados en acciones pedagógicas concretas en las escuelas y redes de enseñanza.





Un sistema consolidado en la región

El panorama colombiano revela un sistema de evaluación a gran escala relativamente consolidado en América Latina, con trayectoria continua, fuerte comparabilidad territorial y amplia disponibilidad de datos.

La evaluación Saber cumple un papel central en el seguimiento del desempeño educativo del país, ofreciendo información relevante tanto para el debate público como para la formulación de políticas.

Los desafíos futuros parecen estar menos ligados a la construcción del sistema evaluativo en sí —que ya presenta bases técnicas sólidas— y más relacionados con la capacidad de transformar estas informaciones en mejoras sostenidas en el aprendizaje de los estudiantes.

Como veremos en los demás países de los textos de esta publicación, las evaluaciones a gran escala ocupan posiciones distintas dentro de los sistemas educativos latinoamericanos. Comprender estas diferencias ayuda a entender no sólo los resultados observados, sino también la capacidad de cada país para utilizar datos educativos como instrumento efectivo de política pública.

PARTE II — SISTEMAS NACIONAIS EN PERSPECTIVA

**Capítulo 8****México: la ausencia como elección y sus efectos sobre el sistema**

A lo largo de las últimas décadas, México llegó a construir uno de los sistemas de evaluación educativa más estructurados de América Latina. Diferentes instrumentos nacionales —como EXCALE, ENLACE y posteriormente PLANEA— permitieron acompañar la evolución de los resultados educativos, producir diagnósticos del sistema y alimentar el debate público sobre la calidad de la educación.

Actualmente, sin embargo, México no posee una evaluación nacional estandarizada de aprendizaje a gran escala con carácter sumativo, es decir, una evaluación capaz de medir de forma comparable lo que los estudiantes están aprendiendo a lo largo del tiempo y entre diferentes regiones del país.



La trayectoria mexicana en evaluación educativa comienza todavía en los años 1980, con la creación del Sistema Nacional de Evaluación. A partir de los años 2000, el país avanzó significativamente en la institucionalización de este campo con la creación del Instituto Nacional para la Evaluación de la Educación (INEE), responsable de desarrollar evaluaciones nacionales y consolidar una cultura de evaluación educativa.

En este periodo, el país contó con evaluaciones nacionales relevantes, entre ellas:

- EXCALE (2005)
- ENLACE (2006)
- PLANEA (2015)

Estas evaluaciones permitían medir el rendimiento de los estudiantes en diferentes etapas de la educación básica, con resultados comparables entre estados y a lo largo del tiempo. Paralelamente, el INEE se consolidó como un organismo técnico relativamente autónomo, responsable de garantizar la calidad metodológica y la transparencia en la divulgación de los datos.

Este arreglo institucional acercaba a México a modelos observados en otros países de la región que desarrollaron sistemas nacionales sólidos de evaluación educativa.

Un cambio de enfoque: de la evaluación estandarizada a la evaluación diagnóstica

Este escenario comenzó a cambiar a partir de la reforma educativa de 2019. En este momento, el INEE fue extinguido y sustituido por la Comisión Nacional para la Mejora Continua de la Educación (Mejoredu), que pasó a tener un mandato más restringido y menor autonomía institucional.

Posteriormente, en 2025, la propia Mejoredu también fue extinguida en el contexto de una reforma más amplia que eliminó diversos organismos autónomos o descentralizados de la administración pública. Como resultado, el país dejó de contar con una institución nacional dedicada a la evaluación educativa con el mismo diseño institucional anterior.

Este proceso acabó rediseñando el papel de la evaluación educativa dentro del sistema mexicano.

Paralelamente a estos cambios institucionales, también hubo una alteración importante en la concepción de evaluación educativa adoptada por el país.

La política educativa asociada al modelo de la Nueva Escuela Mexicana, lanzada en 2022, pasó a privilegiar un enfoque formativo y cualitativo de la evaluación. En este modelo, el objetivo principal no es clasificar o comparar estudiantes por medio de pruebas estandarizadas, sino apoyar el proceso de aprendizaje dentro de la propia escuela.



En este contexto, surgieron los Ejercicios Integradores del Aprendizaje (EIA), instrumentos diagnósticos aplicados al inicio del ciclo escolar. Estos ejercicios se organizan por fases —que agrupan varios años escolares— y buscan ofrecer a los docentes información sobre el nivel de desarrollo de los estudiantes en diferentes campos formativos.

El enfoque es esencialmente pedagógico:1

- **la aplicación es voluntaria,**
- **los propios docentes corrigen las actividades con base en rúbricas,**
- **los resultados se utilizan principalmente dentro de la escuela para orientar el trabajo pedagógico.**

Este modelo tiene coherencia con una visión formativa de la evaluación, cada vez más presente en diversos sistemas educativos.



Las limitaciones del modelo actual

A pesar de sus virtudes pedagógicas, el modelo vigente presenta algunas limitaciones cuando se analiza bajo la perspectiva del acompañamiento del sistema educativo.

Primero, los instrumentos actuales no fueron concebidos para permitir comparaciones entre estados o a lo largo del tiempo. El foco es el diagnóstico local —aula, grupo o escuela— y no la medición del rendimiento del sistema educativo como un todo.

Segundo, los resultados no se publican a nivel nacional o subnacional, permaneciendo principalmente como insumo para el trabajo pedagógico de las escuelas.

Tercero, la responsabilidad por la evaluación pasó a estar concentrada en áreas internas de la Secretaría de Educación Pública, en un arreglo institucional diferente del que existía cuando había organismos especializados dedicados exclusivamente a esta función.

En la práctica, el país pasó de un modelo fuertemente estructurado en evaluaciones nacionales estandarizadas a otro en el que la evaluación tiene un carácter más pedagógico y situado en el contexto de las escuelas.

No hay dudas sobre la importancia de las evaluaciones formativas. En este sentido, el enfoque actualmente adoptado en México posee muchos méritos. No obstante, las evaluaciones de aprendizaje a gran escala, cuando están bien diseñadas e implementadas, cumplen otro papel igualmente fundamental. Son instrumentos imprescindibles para orientar políticas públicas educativas y acompañar la evolución de los sistemas de enseñanza, especialmente en países de América Latina, cuyos sistemas educativos están en un nivel de madurez institucional que hace que este tipo de monitoreo sea particularmente relevante.

El impacto en la producción de información educativa

Este cambio tiene implicaciones importantes para la producción de información sobre el sistema educativo.

Sin evaluaciones nacionales comparables, se vuelve más difícil responder a preguntas que normalmente orientan el diseño de políticas educativas, como por ejemplo:

- ¿Cómo están evolucionando los aprendizajes de los estudiantes a lo largo del tiempo?
- ¿Existen diferencias importantes entre estados o regiones?
- ¿Qué políticas educativas están generando mejores resultados?

Estas preguntas siguen siendo relevantes para cualquier sistema educativo, y diferentes países han buscado equilibrar el uso de evaluaciones formativas a nivel de la escuela con instrumentos nacionales que permitan acompañar tendencias del sistema.

El caso mexicano en el contexto latinoamericano

Cuando se compara con otros países de la región analizados en los capítulos anteriores —como Brasil, Perú o Argentina—, México presenta hoy una situación particular.

Mientras que estos países mantienen evaluaciones nacionales periódicas capaces de producir series históricas y diagnósticos comparables, México atraviesa un periodo de redefinición del papel de la evaluación educativa dentro de su política pública.

Al mismo tiempo, permanece un legado institucional y técnico relevante acumulado a lo largo de décadas. Parte de esta información histórica, incluso, está siendo preservada por organizaciones de la sociedad civil que buscan evitar la pérdida de este patrimonio de datos educativos.

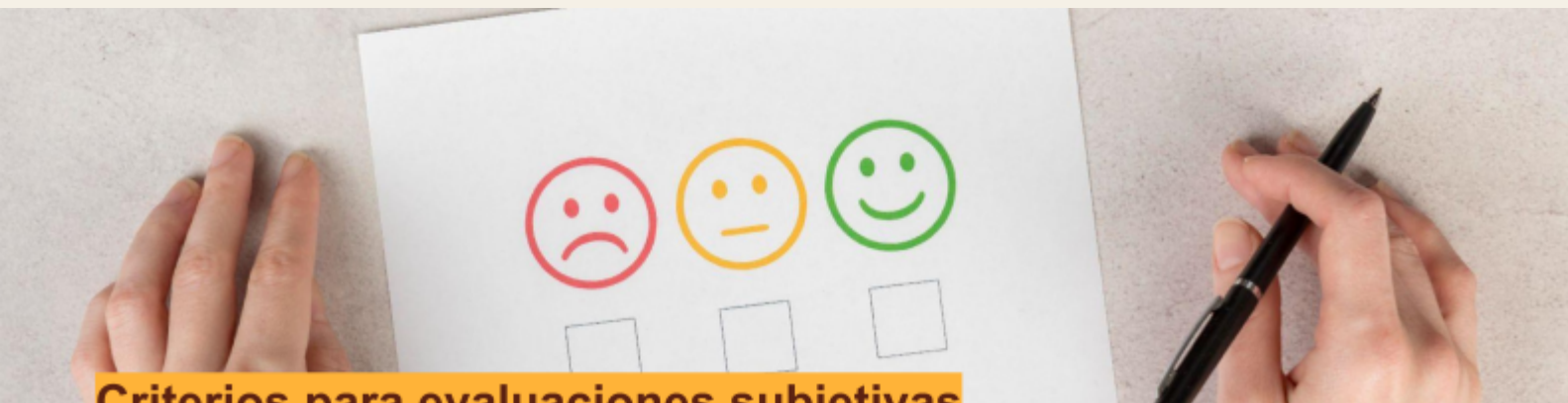


PARTE II – SISTEMAS NACIONAIS EM PERSPECTIVA

Capítulo 9

Un cuadro comparado para América Latina

Item de análise	Brasil		Argentina		Perú	Chile	Colômbia	México
Evaluación	SAEB Alfabetização	SAEB	APRENDER Alfabetização	APRENDER	ENLA	SIMCE	SABER	NÃO HÁ
1. Séries/años evaluados	2º ano do EF	5º e 9º años EF; 3º año de EM	3º año de primário	6º año de primário	2º y 4º do primário; 2º do secundário	4º, 6º y 8º Básica	Educación Básica – Años Iniciales: 3º y 5º años	
2. Calidad técnica (grado de consenso nacional)	Media	Media	Media	Media	Alta	Media	Alta	
3. Periodicidad	Anual	Bienal	No definida	Bienal	Anual (alternados)	Anual (4º) / Bienal	A cada 2 ou 3 anos	
4. Apertura y comparabilidad por entidades subnacionales o regionales	Alta	Alta	Media	Alta	Media	Alta	Alta	
5. Comparabilidad en el tiempo	Media	Alta	Bajo	Alta	Media	Alta	Alta	
6. Transparencia de datos e metodologia	Alta	Alta	Alta	Alta	Alta	Alta	Media	
7. Independencia técnica e institucional del organismo responsable por la medición	Alta	Alta	Media	Media	Media	Media/Alta	Alta	
8. Publicación de resultados	Alta	Alta	Alta	Alta	Alta	Alta	Media	
9. Grado de institucionalización	Media	Alto	Bajo	Alto	Alto	Alto	Media	
10. Uso coherente de tecnologia	Media	Media	Media	Media	Alto	Alto	Media	
11. Años de historial	4 (desde 2021)	Mais de 10 ciclos (1ª aplicação em 1990)	1 (2024)	7 (desde 2016)	18 (2007-2025)	22 (desde 2002)	13 anos (2012-2025)	
12. Resultados inteligibles para públicos no especializados	Alto	Media	Bajo	Media	Alto	Alto	Media	
13. Uso de resultados y nivel de incidencia en políticas públicas	Alto	Alto	Alto	Médio	Médio	Alto	Médio	



Criterios para evaluaciones subjetivas

Calidad técnica (grado de consenso nacional)

Alto consenso: Evaluación ampliamente reconocida por especialistas, academia, gestores y órganos técnicos; metodología estable; críticas concentradas en el uso, no en el instrumento.

Consenso intermedio: Reconocimiento técnico relevante, pero con controversias recurrentes sobre diseño, metodología o interpretación.

Bajo consenso: Cuestionamientos frecuentes sobre validez, confiabilidad o adecuación técnica del instrumento.

Apertura y comparabilidad por entidades subnacionales o regionales

Alta: Resultados comparables entre estados/municipios/regiones, con recortes públicos y metodológicamente consistentes.

Media: Comparabilidad posible solo para algunos niveles o con limitaciones relevantes.

Baja o inexistente: Resultados agregados nacionalmente, sin posibilidad real de comparación subnacional.

Comparabilidad en el tiempo

Alta: Escala y metodología estables; series históricas continuas; cambios raros y bien documentados.

Media: Comparabilidad posible, pero con rupturas metodológicas puntuales.

Baja: Cambios frecuentes que dificultan la lectura de tendencia.

Transparencia de datos y metodología

Alta: Microdatos, informes técnicos y documentación metodológica públicos y accesibles.

Media: Informes disponibles, pero con limitaciones de acceso a datos o detalles metodológicos.

Baja: Poca información pública sobre metodología o datos.

Independencia técnica e institucional del órgano responsable de la medición

Alta: Órgano técnico con autonomía institucional, reglas claras de gobernanza y protección contra interferencia político-partidaria.

Media: Órgano ligado al gobierno, pero con tradición técnica consolidada y algún grado de protección.

Baja: Fuerte dependencia política, con cambios frecuentes asociados a ciclos de gobierno.

Publicidad de los resultados

Alta: Divulgación amplia, regular y accesible, con informes públicos y comunicación activa.

Media: Divulgación pública, pero con alcance restringido o poca sistematicidad.

Baja: Divulgación limitada o restringida a públicos técnicos.

Grado de institucionalización

Alta: Evaluación prevista en ley o normativa estable; continuidad entre gobiernos.

Media: Evaluación recurrente, pero basada en decisiones administrativas.

Baja: Evaluación puntual o sujeta a discontinuidad política.



Uso coherente de tecnología

Alto:

- La tecnología se usa de forma alineada a la finalidad de la evaluación (monitoreo, comparabilidad, política pública).
- Contribuye claramente a: reducción de costos; mayor escala o cobertura; ciclos evaluativos más cortos; mejor comunicación de los resultados.
- No compromete: comparabilidad en el tiempo; equidad de acceso; validez del instrumento.
- La tecnología está subordinada al diseño pedagógico y técnico, no lo contrario.

Medio:

- La tecnología se utiliza, pero con ganancias limitadas o concentradas en una sola dimensión (ej.: logística); o aún en fase de transición/piloto.
- Hay beneficios claros, pero también: restricciones de acceso; limitaciones técnicas; o dependencia de infraestructura desigual.
- El uso tecnológico no distorsiona el objetivo, pero todavía no lo potencializa plenamente.

Bajo:

- Uso tecnológico: inexistente; o desarticulado de la finalidad de la evaluación; o adoptado más por moda que por necesidad técnica.
- Puede generar: aumento de desigualdades; pérdida de comparabilidad; fragilidad técnica u operacional.
- La tecnología se sobrepone al diseño evaluativo, en lugar de servirlo.

Resultados inteligibles para públicos no especializados

Alta: Resultados comunicados en métricas intuitivas (porcentajes, niveles claros), comprensibles para el público general.

Media: Resultados comprensibles con mediación técnica.

Baja: Resultados presentados solo en escalas técnicas o puntuaciones abstractas.

Uso de los resultados y nivel de incidencia en políticas públicas

Alto: Resultados claramente utilizados para diseño, ajuste o monitoreo de políticas públicas.

Medio: Uso parcial o indirecto, sin vínculo sistemático con decisiones.

Bajo: Resultados poco utilizados o meramente informativos.



Conclusión

El futuro de las evaluaciones a gran escala

A lo largo de esta publicación, buscamos discutir el papel de las evaluaciones a gran escala en América Latina a partir de una perspectiva menos polarizada y más anclada en la realidad de los sistemas educativos. Estos textos fueron contruidos en diálogo con Karina Stocovaz, Gerente Sr. del Instituto Natura para los países de habla hispana, cuya contribución fue fundamental para ampliar la visión regional.

Las evaluaciones, como vimos, están lejos de ser instrumentos perfectos. No capturan toda la complejidad de la educación, pueden inducir distorsiones cuando se utilizan mal y no sustituyen políticas públicas bien diseñadas. Al mismo tiempo, su ausencia compromete dimensiones centrales del funcionamiento de los sistemas educativos: la capacidad de diagnóstico, la transparencia y el propio sentido de responsabilización pública.



La experiencia latinoamericana revela un escenario diverso, con países en diferentes etapas de desarrollo de sus sistemas evaluativos. Esta heterogeneidad no debe verse sólo como un desafío, sino también como una oportunidad.

Hay espacio concreto para el aprendizaje entre países, la circulación de buenas prácticas y la aceleración de trayectorias a partir de experiencias ya consolidadas en la región.

Avanzar, en este contexto, no significa simplemente expandir evaluaciones, sino mejorar continuamente su diseño, su uso y su integración con las políticas educativas. Significa también reconocer sus límites, evitando tanto la sobrevaloración como el descarte precipitado.

Si están bien construidas y se utilizan con responsabilidad, las evaluaciones a gran escala siguen siendo uno de los pocos instrumentos capaces de sostener, al mismo tiempo, visión sistémica, coordinación y presión por la mejora. No resuelven, por sí solas, los desafíos educativos de la región, pero son parte indispensable de cualquier estrategia que pretenda enfrentarlos con seriedad y consistencia.



instituto
natura



AVON