

Edição 2026



# Avaliações de Aprendizagem em Larga Escala na América Latina

Da mensuração à transformação  
dos sistemas educacionais

Organização: **David Saad**, Diretor-Presidente  
do Instituto Natura para a América Latina

# Sumário

## 3 **Apresentação**

### PARTE I – O PAPEL DAS AVALIAÇÕES

## 4 **Capítulo 1** – Por que medir aprendizagem importa

## 12 **Capítulo 2** – Qualquer avaliação de aprendizagem em escala serve?

### PARTE II – SISTEMAS NACIONAIS EM PERSPECTIVA

## 18 **Capítulo 3** – **Brasil:** robustez e complementaridade

## 23 **Capítulo 4** – **Argentina:** tradição e desafios de consolidação

## 28 **Capítulo 5** – **Peru:** institucionalização com tensões metodológicas

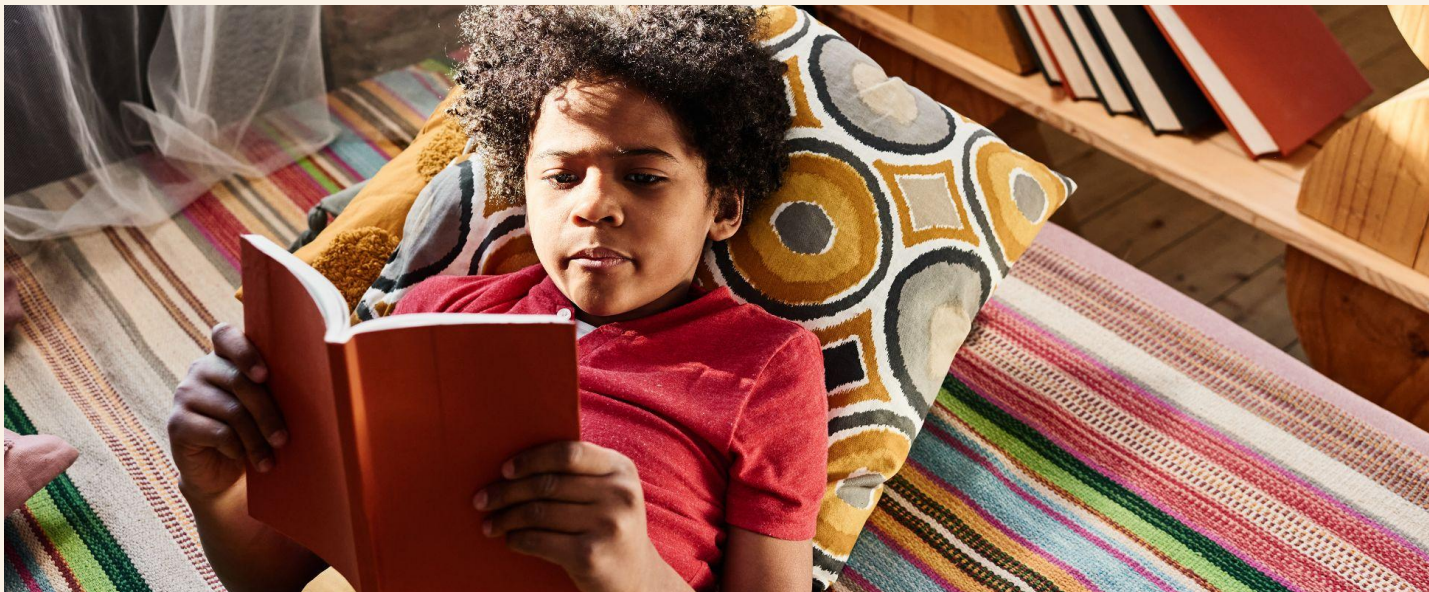
## 33 **Capítulo 6** – **Chile:** maturidade, crise e reinvenção

## 37 **Capítulo 7** – **Colômbia:** sistema estruturado e consistente

## 42 **Capítulo 8** – **México:** da referência à ausência de avaliação nacional

## 48 **Capítulo 9** – Um quadro comparado para a América Latina

## 53 **Conclusão** – O futuro das avaliações em larga escala



## APRESENTAÇÃO

Este e-book reúne, organiza e aprofunda uma série de artigos dedicados ao papel das avaliações de aprendizagem em larga escala nos sistemas educacionais da América Latina.

Os textos, escritos em parceria com Karina Stocovaz, Superintendente do Instituto Natura para os países de língua hispânica, partem de uma premissa central: em contextos marcados por desigualdades educacionais profundas, não é possível melhorar aquilo que não se mede de forma consistente, comparável e pública.

Ao longo das últimas décadas, avaliações em larga escala deixaram de ser instrumentos periféricos para ocupar um lugar central no debate educacional. Elas passaram a cumprir a função estratégica de tornar visível o nível de aprendizagem dos estudantes, orientar decisões de política pública e fortalecer o compromisso político com resultados.

Ao mesmo tempo, essas avaliações também carregam limites e riscos reais — desde custos elevados até distorções de comportamento e simplificação da realidade educacional.

**O objetivo desta publicação é justamente explorar essa tensão, reconhecer os limites sem perder de vista o papel estruturante das avaliações.**

Este e-book busca responder a uma pergunta central: como diferentes países utilizam, ou deixam de utilizar, avaliações em larga escala para melhorar seus sistemas educacionais?

## PARTE I — O PAPEL DAS AVALIAÇÕES

**Capítulo 1****Por que medir aprendizagem importa**

Para que sistemas educacionais — especialmente na América Latina — consigam avançar de forma consistente, é fundamental dispor de mecanismos que tornem visível o que os estudantes estão aprendendo, permitam acompanhar a evolução ao longo do tempo e criem condições para ajustes de políticas públicas.

**Avaliações em escala não são políticas educacionais em si, nem instrumentos pedagógicos de sala de aula.**

**Elas cumprem outra função: ajudam a estruturar o debate público, orientam decisões em nível sistêmico e fortalecem o compromisso político com a melhoria da aprendizagem.**

Apesar de muitas experiências que demonstram a potência e a importância das avaliações de aprendizagem em escala, também existem limitações relevantes associadas a esse tipo de instrumento. É sobre esses limites que este texto trata — e sobre por que, mesmo reconhecendo-os, as avaliações em escala continuam sendo fundamentais

Antes, porém, é importante delimitar o escopo da discussão.



Existem usos inadequados das avaliações em larga escala. Há atores que se apropriam desses instrumentos sem compromisso com a melhoria da educação, utilizando seus resultados para fins que pouco têm a ver com aprendizagem. Da mesma forma, há rejeições às avaliações que não decorrem de limitações técnicas ou pedagógicas, mas do desconforto com a ideia de monitoramento, comparação e responsabilização.

Essas posturas não partem de um olhar orientado à melhoria dos sistemas educativos. Por isso, não são o foco desta análise. O debate que segue parte do pressuposto de que o objetivo é, de fato, melhorar a aprendizagem — e fazê-lo com equidade.

## **O custo das avaliações em larga escala**

Uma crítica recorrente diz respeito ao custo. Avaliações de aprendizagem em escala são caras, especialmente quando bem desenhadas e implementadas como devem ser, e mais ainda quando adotam desenhos censitários. Esse é um ponto que merece ser reconhecido sem rodeios.

**Ao mesmo tempo, a experiência acumulada em diversos países mostra que essas avaliações podem funcionar como um importante driver de melhoria dos sistemas educacionais, das políticas públicas e, como consequência, dos resultados de aprendizagem.**

Sob essa perspectiva, mesmo com custos elevados, o investimento tende a se justificar.

Além disso, existem caminhos para mitigar esses custos. Um deles diz respeito ao desenho da avaliação. Quando o objetivo central é monitoramento e geração de compromisso político, avaliações amostrais podem cumprir bem esse papel. Se um país ou uma entidade não dispõe de recursos para uma avaliação censitária, é preferível realizar uma avaliação amostral do que simplesmente não avaliar.

Em países federativos, também existem arranjos institucionais alternativos. As avaliações não precisam ser exclusivamente nacionais. Podem existir avaliações subnacionais, desde que parametrizadas de modo a garantir comparabilidade. O Indicador Criança Alfabetizada no Brasil, é um exemplo desse tipo de arranjo.

Outro caminho relevante é o fortalecimento de entidades técnicas especializadas em avaliação em larga escala. A atuação de parceiros técnicos contratados pelo poder público pode gerar ganhos de escala, eficiência e redução de custos, especialmente quando esses atores operam em múltiplas redes ou territórios.

Por fim, o uso de tecnologia abre novas possibilidades. Avaliações de fluência com suporte de IA, amplamente utilizadas em alfabetização, são um exemplo já consolidado. Outras metodologias mais recentes, como a avaliação ROAR (Rapid Online Reading Assessment), apontam para caminhos em que a mensuração da aprendizagem pode se tornar significativamente mais barata.

## A utilidade para o professor e para a sala de aula

Outra crítica frequente é que avaliações em larga escala pouco ajudam o professor no trabalho cotidiano em sala de aula. Em geral, os resultados demoram a ser divulgados, são apresentados de forma agregada e não permitem diagnósticos individualizados dos estudantes.

Essa crítica é, em parte, verdadeira. Avaliações em escala, na maioria das vezes, não substituem avaliações diagnósticas e individualizadas, que são fundamentais para o trabalho pedagógico direto com o aluno. Idealmente, os sistemas educacionais deveriam contar com ambos os tipos de instrumento.





Dependendo do desenho, no entanto, avaliações em escala podem sim gerar informações úteis para escolas e professores — tema que será aprofundado no próximo capítulo. Ainda assim, o ponto central aqui é outro: as avaliações não precisam cumprir todas as funções ao mesmo tempo.

O fato de uma avaliação em escala não servir diretamente como instrumento diagnóstico não é um argumento contra sua realização. Seu papel principal, neste debate, é o monitoramento sistêmico, o fortalecimento do compromisso político e a orientação de políticas públicas. Mesmo sem uso direto em sala de aula, ela pode fazer toda a diferença para o funcionamento do sistema como um todo.

### **O risco de estigmatização**

A estigmatização é um risco real associado às avaliações em larga escala, especialmente quando seus resultados são desagregados em níveis subnacionais, e quando possível, por escola. Esses dados podem ser usados para rotular estudantes, escolas, professores, gestores ou redes, gerando efeitos negativos que pouco contribuem para a melhoria da educação.

Esse risco, no entanto, não está na avaliação em si, mas no mau uso dos seus resultados. Utilizações com viés punitivo ou exclusivamente negativo tendem a produzir mais efeitos adversos do que ganhos. A ideia de que expor resultados ruins, por si só, levará à melhoria da educação parte de uma visão simplista e pouco realista sobre o funcionamento dos sistemas educacionais.

**Avaliações de aprendizagem não deveriam ser utilizadas para estigmatizar. Esse risco pode — e deve — ser enfrentado no desenho e na implementação da avaliação.**

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



## A simplificação da realidade educacional

Avaliações em larga escala são, por definição, uma simplificação da realidade. Elas não capturam o desenvolvimento integral de crianças e jovens. Em geral, concentram-se em língua e matemática, às vezes em ciências, deixando de fora diversas dimensões centrais da formação humana.

Essa limitação precisa ser reconhecida. O problema surge quando os resultados da avaliação passam a ser tratados como objetivo final, e não como uma mensuração aproximada da realidade. Melhorar o indicador não é sinônimo de melhorar a educação.

Por outro lado, como todo indicador, avaliações em escala são úteis mesmo sendo imperfeitas. Há evidências consistentes de correlação entre os resultados dessas avaliações e dimensões mais amplas do desenvolvimento dos estudantes. Além disso, elas medem conhecimentos básicos que são condições necessárias, ainda que não suficientes, para o desenvolvimento integral.

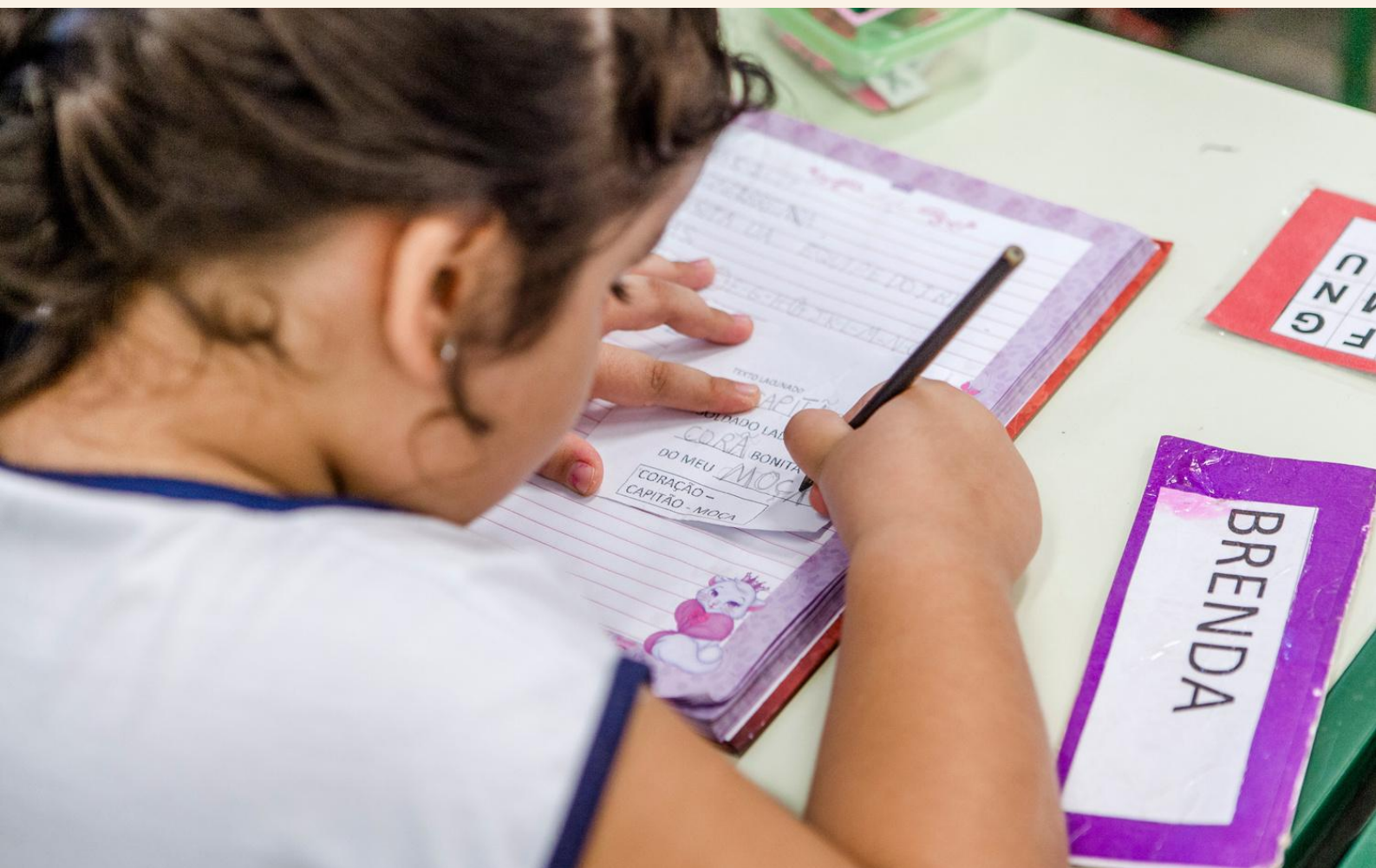
No contexto da América Latina, onde déficits nesses conhecimentos ainda são elevados, faz muito sentido mensurá-los. O indicador deve ser entendido como uma leitura importante, mas simplificada da realidade — e não como o estado ideal que se deseja alcançar.

## Ensino voltado para a prova e distorções de comportamento

Esse ponto está diretamente relacionado ao anterior. Quando sistemas educacionais passam a entender que o objetivo do processo educativo é ir bem na avaliação, surgem distorções. A avaliação deixa de ser instrumento e passa a ser fim.

Essas distorções nem sempre decorrem de decisões explícitas de “treinar para a prova”. Muitas vezes são resultado de incentivos mal desenhados dentro do ecossistema educacional. O risco não está em utilizar avaliações em escala, mas em confundir seu papel.

Avaliações em escala medem um mínimo necessário, com correlação com o desenvolvimento integral, mas não esgotam os objetivos da educação. É fundamental que o desenho e a implementação dessas avaliações incorporem mecanismos que reduzam esse tipo de efeito.



## Ideologização da avaliação de aprendizagem

Por fim, há a crítica da ideologização. Avaliações em larga escala podem suscitar debates filosóficos legítimos sobre mérito, justiça, equidade e visões de mundo. Essas discussões são importantes e devem acontecer.

Avaliações de aprendizagem em escala têm limites e riscos reais. Elas não capturam toda a complexidade da educação, podem gerar distorções quando mal utilizadas e levantam debates legítimos sobre custo, estigmatização e sentido pedagógico. Reconhecer essas limitações é essencial para qualificar o debate — e não para evitá-lo.



Nesta mini-publicação, discuto esses limites de forma direta e explico por que, ainda assim, as avaliações em larga escala seguem sendo fundamentais para monitorar sistemas educacionais, orientar políticas públicas e fortalecer o compromisso político com a melhoria da aprendizagem, especialmente em contextos marcados por desigualdades como o da América Latina.

No entanto, quando o foco se desloca para o objetivo de melhorar a aprendizagem com equidade, a avaliação se mostra uma ferramenta potente. É difícil sustentar, em termos substantivos, a rejeição a objetivos como alfabetizar todas as crianças na idade certa ou garantir que jovens concluam o ensino médio com um nível básico de aprendizagem.

A experiência internacional mostra que avaliações em larga escala são adotadas em países com orientações ideológicas muito distintas. Chile e Brasil, por exemplo, sob governos de diferentes espectros políticos, mantiveram e fortaleceram seus sistemas de avaliação. Isso sugere que esse instrumento pode — e deveria — ultrapassar disputas partidárias.

**A avaliação em escala não precisa dar conta de todos os objetivos da educação. O fato de ela cumprir bem o papel de monitoramento e compromisso político já pode ser suficiente, especialmente quando não há soluções viáveis para que ela assuma outros usos.**



## **Considerações finais**

Reconhecer que avaliações de aprendizagem em escala têm limitações é fundamental. Esses limites existem, são reais e precisam ser levados a sério. Ao mesmo tempo, muitos desses riscos podem ser enfrentados por meio de escolhas cuidadosas no desenho e na implementação das avaliações.

Mesmo com todas essas limitações, o poder das avaliações em larga escala em termos de monitoramento, ajuste de políticas públicas, fortalecimento do compromisso político e envolvimento de diferentes atores da sociedade é grande o suficiente para justificar sua realização.

No próximo capítulo, a discussão avança para outro nível: quais são as características das avaliações de aprendizagem em escala que ajudam a mitigar esses riscos e aumentam as chances de que elas sejam efetivamente usadas para melhorar a aprendizagem com equidade.

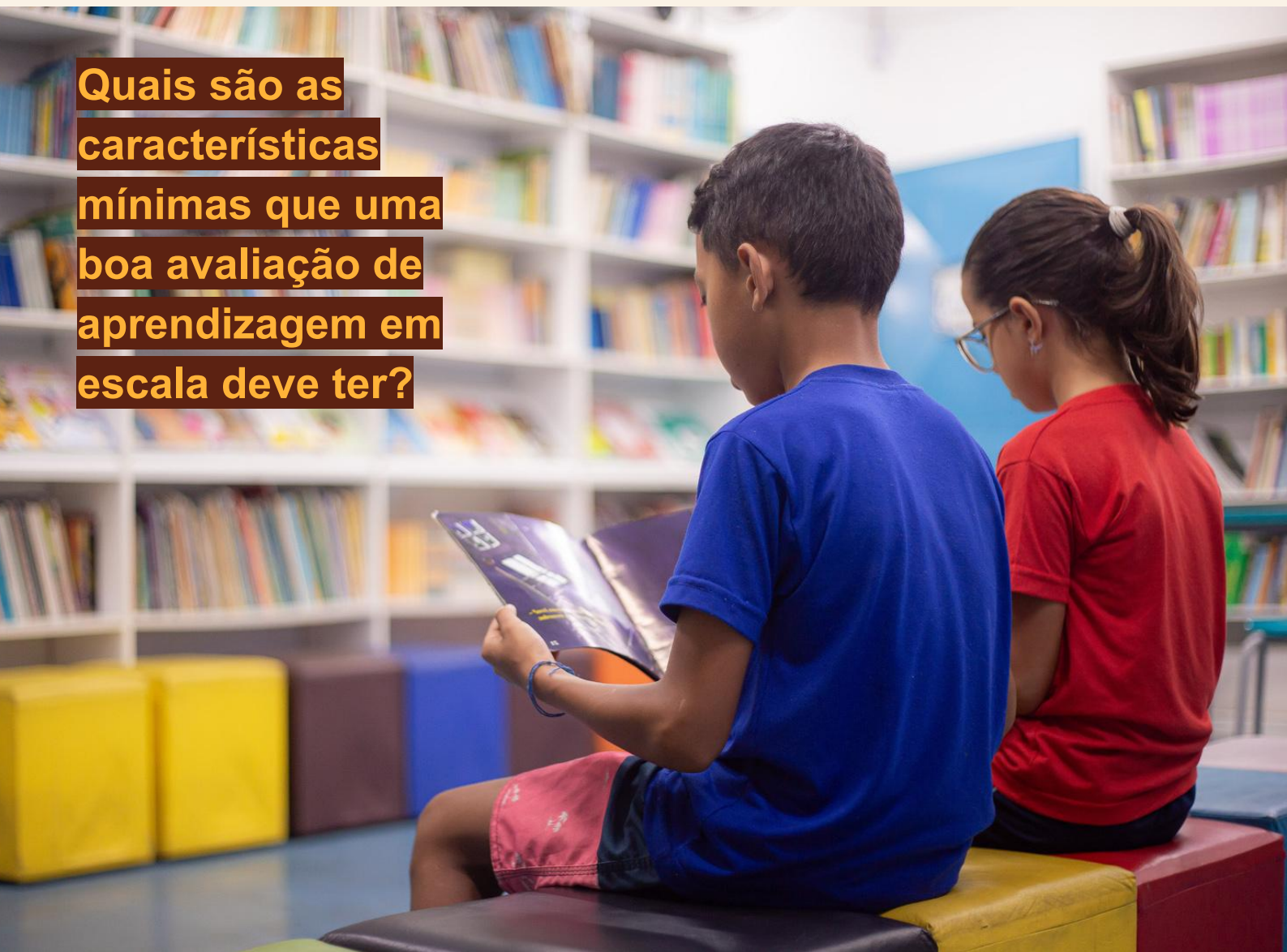
## PARTE I – O PAPEL DAS AVALIAÇÕES

**Capítulo 2****Qualquer avaliação de aprendizagem em escala serve?**

Uma vez que consideramos a avaliação de aprendizagem em escala fundamental para gerar e sustentar o compromisso político — além de permitir a revisão e o ajuste de políticas públicas —, e mesmo reconhecendo suas limitações, gostaria de me deter, neste texto, na discussão sobre quais características essas avaliações precisam ter para cumprir adequadamente o seu papel.

Vale lembrar que avaliações em escala podem cumprir diferentes funções. Aqui, no entanto, refiro-me especificamente ao que considero seu papel principal: apoiar a tomada de decisão em políticas públicas e fortalecer o compromisso político com a melhoria da aprendizagem.

**Quais são as características mínimas que uma boa avaliação de aprendizagem em escala deve ter?**



## Objetivo

A avaliação de aprendizagem em escala deve existir, antes de tudo, com o objetivo de melhorar algum resultado educacional — em geral por meio da implementação ou da revisão de políticas públicas. Quando não há a intenção real de atuar sobre os resultados, ou, ao menos, de chamar atenção para a necessidade de fazê-lo, a avaliação se torna inútil e, por vezes, contraproducente.

Curiosamente, o raciocínio inverso também costuma ser verdadeiro: na ausência de avaliações em escala com as características discutidas aqui, torna-se difícil acreditar que o ente público esteja genuinamente comprometido em melhorar os resultados educacionais em questão.



## Qualidade técnica

Se desejamos mensurar, por exemplo, o nível de alfabetização na idade certa em um país, estado ou município, é fundamental que o instrumento de avaliação seja tecnicamente adequado — isto é, que efetivamente mensure o que se propõe a medir.

Embora isso pareça óbvio, alcançar esse objetivo exige um esforço significativo: definição clara de critérios, construção de progressões e elaboração de questões (ou outros instrumentos) que sejam, de fato, capazes de avaliar o nível de aprendizagem pretendido.

## Comparabilidade no espaço e no tempo

Se o objetivo é gerar compromisso político e, como consequência, ajustes em políticas públicas, é fundamental que as avaliações sejam comparáveis ao longo do tempo. A avaliação perde grande parte de seu valor — ou pode até se tornar enviesada, por vezes com objetivos políticos — quando não permite a leitura consistente de avanços ou retrocessos ao longo dos anos.



Também é altamente desejável que haja comparabilidade no espaço, entre estados, municípios ou regiões. Não se trata de criar rankings, mas de compreender onde as políticas têm funcionado melhor ou pior, permitindo identificar o que funciona, o que não funciona e onde houve — ou não — compromisso político.

Não entrarei aqui em detalhes técnicos, mas vale destacar a importância tanto da construção dos itens (por exemplo, com o uso da metodologia TRI) quanto do desenho amostral e de outros aspectos metodológicos. Mesmo avaliações censitárias nunca alcançam 100% de participação, e garantir que a população avaliada não esteja distorcida é um ponto central.

## Periodicidade

Para os objetivos mencionados, a repetição da avaliação com uma periodicidade adequada é fundamental. Entendo que a periodicidade máxima deveria ser de dois anos.

Reconheço que existe um debate segundo o qual mudanças nos resultados educacionais levam muito tempo e, portanto, avaliações em intervalos curtos seriam pouco informativas. No entanto, tenho observado a importância da periodicidade tanto para sustentar o compromisso político quanto para captar avanços mais rápidos do que o esperado, especialmente nos anos iniciais.

Um exemplo é o Indicador Criança Alfabetizada (ICA), implementado no Brasil com periodicidade anual, que vem demonstrando uma tendência de melhora. Caso fosse aplicado a cada cinco anos, esses avanços não seriam percebidos, e tampouco ocorreriam, com a mesma intensidade, os ajustes nas políticas de alfabetização em regime de colaboração ou o reforço positivo aos gestores que estão avançando.



## **Confiabilidade e transparência**

Se o objetivo é alterar políticas públicas para alcançar melhores resultados de aprendizagem — e se a implementação final dessas mudanças ocorre nos sistemas locais e nas escolas —, é fundamental que haja confiança nos resultados.

Isso implica assegurar que os dados sejam reais, que não haja objetivos escusos e que não existam falhas metodológicas ocultas. Confiabilidade e transparência são condições necessárias para que os resultados sejam apropriados pelos diferentes atores envolvidos.

## Publicização dos resultados

Este é um ponto básico. Realizar avaliações de aprendizagem em escala e restringir o acesso às informações a um pequeno grupo de técnicos reduz significativamente sua relevância para os objetivos aqui discutidos.

Mesmo considerando os riscos de estigmatização, ranqueamentos e uso inadequado dos dados — que devem ser mitigados tanto quanto possível —, entendo que, para os fins descritos neste texto, a publicização dos resultados deve ser a mais ampla possível.

Também é importante que os resultados sejam compreensíveis para um público mais amplo, e não apenas para especialistas. Retomando o exemplo do ICA: divulgar que a média nacional foi de 560 pontos (número ilustrativo) diz pouco para a maioria das pessoas. Informar que, em 2023, 59,2% das crianças estavam alfabetizadas na idade certa, comunica de forma muito mais clara o significado do resultado.



## Uso adequado

Por fim, como qualquer instrumento, avaliações de aprendizagem em escala podem ser mal utilizadas. Podem servir apenas para punição, ser instrumentalizadas em disputas políticas ou contribuir para a estigmatização de determinadas regiões.

É necessário trabalhar para mitigar esses riscos. Ainda assim, como argumentei no capítulo anterior, mesmo considerando essas possibilidades de mau uso, entendo que as avaliações de aprendizagem em escala são fundamentais para o avanço da qualidade da educação nos países latino-americanos.



## Nota final sobre terminologia

Nos textos sobre avaliação de aprendizagem em escala, utilizo o termo **avaliação de aprendizagem** para me referir ao que parte da literatura especializada denomina avaliação de desempenho. Tenho clareza dessa distinção conceitual. A opção pelo termo aprendizagem foi deliberada, por considerá-lo mais acessível e menos confuso para o público mais amplo com o qual dialogo aqui, especialmente no debate sobre políticas públicas e sistemas educacionais.

Isso não significa confundir desempenho com aprendizagem, nem atribuir às avaliações externas um papel que elas não têm. Ao longo dos textos, procuro deixar claro que essas avaliações observam desempenho em tarefas estruturadas e produzem evidências que permitem inferências sobre aprendizagem, sem jamais substituir o acompanhamento pedagógico realizado no cotidiano da sala de aula.

A distinção conceitual é importante e exatamente por isso vale explicitá-la.

## PARTE II – SISTEMAS NACIONAIS EM PERSPECTIVA

**Capítulo 3****Brasil: robustez e complementaridade**

Como disse, avaliações de aprendizagem em escala são instrumentos centrais para o monitoramento dos resultados educacionais, para a geração de compromisso político e para o ajuste de políticas públicas — ainda que seja fundamental reconhecer seus limites e riscos.

A partir de agora, avanço um passo adicional. Junto-me a Karina Stocovaz, Superintendente do Instituto Natura para os países de língua hispânica, para realizar uma análise comparativa dos países onde o Instituto Natura atua, tendo como foco as avaliações de aprendizagem em escala.

A análise se baseia em um conjunto de critérios para avaliação de instrumentos, incluindo qualidade técnica, comparabilidade, transparência, institucionalização, uso coerente de tecnologia e incidência em políticas públicas. Ao final dos capítulos, apresentaremos as análises em formato de tabela, de maneira mais objetiva, o que permitirá algum tipo de comparabilidade dos diferentes países.



Começamos com um panorama das avaliações nacionais de aprendizagem em escala existentes hoje no Brasil, observando suas principais características, usos e graus de consolidação institucional.

Atualmente, o Brasil conta com duas avaliações nacionais recorrentes e em larga escala:

- o SAEB, voltado às etapas finais do Ensino Fundamental e ao Ensino Médio;
- o SAEB Alfabetização, focado nos anos iniciais, especialmente no 2º ano do Ensino Fundamental.

Embora façam parte de um mesmo sistema, essas avaliações cumprem papéis distintos e apresentam níveis diferentes de maturidade institucional.

### **SAEB Alfabetização: foco recente, alto uso em política pública**

O SAEB Alfabetização é uma avaliação relativamente recente, iniciada em 2021, voltada à mensuração da alfabetização inicial, com foco em leitura e escrita ao final do 2º ano do Ensino Fundamental.

**Do ponto de vista técnico, há um certo consenso em relação à sua importância e à sua validade como instrumento até aqui, no entanto, há um debate muito forte entre os especialistas sobre a necessidade de se revisar o modelo da avaliação para que se alinhe a inovações metodológicas de avaliações internacionais.**

**A transparência é alta, com ampla disponibilização de microdados, relatórios e documentação metodológica.**

A avaliação tem periodicidade anual, característica que a torna especialmente adequada para o monitoramento contínuo das políticas de alfabetização. Seus resultados são amplamente comparáveis entre estados e municípios e vêm sendo divulgados de forma clara, com indicadores facilmente compreensíveis para públicos não especializados — como o percentual de crianças alfabetizadas na idade certa.

Por outro lado, por ser um instrumento recente, o SAEB Alfabetização ainda possui histórico curto, o que limita análises de longo prazo. Seu grau de institucionalização pode ser considerado intermediário: embora esteja plenamente integrado à política nacional de alfabetização, ainda se encontra em processo de consolidação como política de Estado.

Um ponto de destaque é o uso efetivo de seus resultados em políticas públicas, especialmente no regime de colaboração entre União, estados e municípios, o que reforça seu papel como instrumento de monitoramento e coordenação federativa.

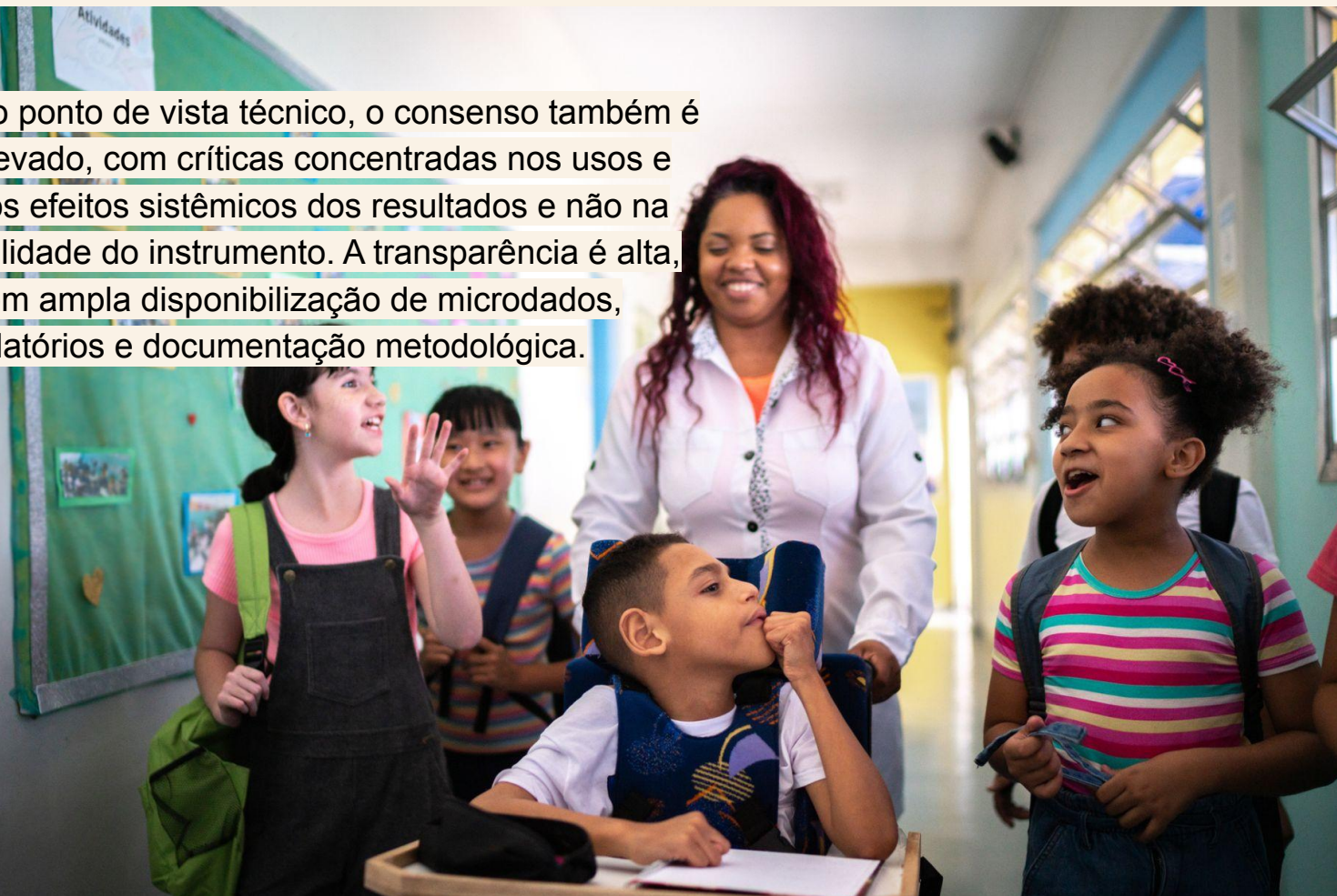


### **SAEB: trajetória longa e papel estruturante no sistema educacional**

O SAEB, por sua vez, é uma das avaliações em larga escala mais consolidadas da região. Aplicado desde 1990, abrange 5º e o 9º anos do Ensino Fundamental e o 3º ano do Ensino Médio, com foco principal em Língua Portuguesa e Matemática.

Trata-se de uma avaliação com alto grau de institucionalização, continuidade ao longo de diferentes governos e por uma extensa série histórica, o que permite leituras consistentes de tendências ao longo do tempo. A comparabilidade territorial é elevada, com ampla desagregação dos resultados por estados, municípios e redes.

Do ponto de vista técnico, o consenso também é elevado, com críticas concentradas nos usos e nos efeitos sistêmicos dos resultados e não na validade do instrumento. A transparência é alta, com ampla disponibilização de microdados, relatórios e documentação metodológica.



Os resultados do SAEB são amplamente divulgados e utilizados em políticas públicas, inclusive em instrumentos como o IDEB, que combina escalas de proficiência em uma métrica mais simples com taxas de aprovação.

Por outro lado, diferentemente do SAEB Alfabetização, os resultados do SAEB, em sua forma original, exigem maior mediação técnica para serem compreendidos pelo público geral, o que limita sua leitura direta fora de círculos especializados.

## **Duas avaliações, papéis complementares**

O panorama brasileiro revela um sistema de avaliação em escala relativamente robusto, com instrumentos que se complementam:

- O SAEB, mais consolidado, cumpre um papel estruturante de monitoramento sistêmico ao longo do tempo;
- O SAEB Alfabetização, mais recente, tem se mostrado particularmente potente para induzir coordenação, foco e ajustes rápidos em políticas de alfabetização.

Ambas apresentam alto grau de uso em políticas públicas, ampla publicidade e forte comparabilidade territorial. As diferenças estão sobretudo no tempo de maturação institucional, na extensão da série histórica e na forma como os resultados dialogam com o público não especializado.

Esse conjunto de características ajuda a explicar por que as avaliações em escala ocupam hoje um lugar central no debate educacional brasileiro — e também aponta para os desafios que permanecem para seu aperfeiçoamento.



## PARTE II – SISTEMAS NACIONAIS EM PERSPECTIVA

**Capítulo 4****Argentina: tradição e desafios de consolidação**

A Argentina conta atualmente com duas avaliações nacionais em larga escala na educação básica:

- Aprender, aplicada principalmente no 6º ano da educação primária;
- Aprender Alfabetização, voltada ao 3º ano e inserida no marco do Plano Nacional de Alfabetização.

Embora façam parte do mesmo sistema, elas se encontram em estágios distintos de maturidade institucional e de consolidação como instrumentos estruturantes de política pública.

## Aprender: tradição avaliativa e debates persistentes

A avaliação Aprender possui trajetória relativamente consolidada desde 2016, com possibilidade de comparação com a edição anterior (Operativo Nacional de Evaluación, de 2013). A Argentina, aliás, acumula experiência avaliativa desde a década de 1990, o que confere densidade técnica ao sistema.

Aplicada de forma censal no 6º ano da educação primária — com alternância para a educação secundária —, a avaliação permite ampla desagregação por jurisdição, setor de gestão e grupos populacionais. Essa característica é particularmente relevante em um país federal como a Argentina, onde as províncias têm papel central na gestão educacional.



A metodologia tem buscado manter estabilidade ao longo do tempo, o que possibilita a leitura de tendências. Ainda assim, o instrumento é objeto de debates recorrentes. Parte das discussões gira em torno da capacidade da avaliação de captar processos mais amplos de aprendizagem; outra parte diz respeito à demanda por divulgação de resultados individualizados por aluno ou escola — algo que não corresponde ao desenho metodológico da prova, construída para leitura sistêmica e não para mensuração individual.



A transparência dos dados é alcançada por meio da publicação de relatórios nacionais e jurisdicionais, microdados (com preservação da confidencialidade), notas metodológicas e plataformas dinâmicas de consulta. Os resultados são amplamente divulgados e frequentemente repercutidos no debate educacional e na mídia.

Ao mesmo tempo, a estrutura institucional responsável pela avaliação está vinculada diretamente à Secretaria de Educação. Embora exista tradição técnica consolidada, não há blindagem institucional plena frente a mudanças de governo, o que poderia gerar instabilidades ao longo do tempo.

No que se refere ao uso em políticas públicas, os resultados são utilizados para identificar escolas que demandam apoio e para fundamentar iniciativas de reforço de aprendizagens. Contudo, esse uso não se apresenta de forma plenamente sistemática ou estruturante no ciclo de formulação de políticas nacionais.

## Aprender Alfabetização: uma linha de base recente

A Aprender Alfabetização foi aplicada pela primeira vez em 2024, com foco no 3º ano da educação primária e inserida no Plano Nacional de Alfabetização.

Trata-se de uma avaliação nova, ainda em fase inicial de consolidação. Sua aplicação teve como objetivo estabelecer uma linha de base para a política nacional de alfabetização, com previsão de nova medição em 2026 para acompanhamento dos avanços.

Por ter introduzido mudanças metodológicas relevantes em relação à avaliação anterior aplicada no 3º ano, não há comparabilidade direta com edições passadas. Isso limita, por ora, a leitura de tendências no tempo.

A avaliação foi amostral, com representatividade por jurisdição, setor de gestão, nível socioeconômico e sexo, o que permite análises relevantes, ainda que com menor grau de detalhamento territorial do que uma aplicação censal.

Para garantir a transparência, foram divulgados relatórios nacionais e jurisdicionais, recomendações pedagógicas, questionários de contexto e bases de dados públicas.

Um dos principais debates concentrou-se na definição do ponto de corte que caracterizaria uma criança como alfabetizada. Foram divulgados múltiplos níveis de desempenho, mas a ausência de um marco público claro que sintetize a informação em uma métrica simples dificulta a comunicação dos resultados para públicos não especializados.



Por outro lado, o uso da avaliação como linha de base para a política nacional de alfabetização é um elemento relevante. A própria previsão de nova aplicação com o objetivo explícito de medir avanços associados à política indica uma intenção de vincular avaliação e ação pública de maneira mais direta.

Ainda assim, trata-se de um instrumento com histórico muito curto e com continuidade institucional ainda em processo de definição.

### **Um sistema com bases sólidas, mas desafios de consolidação**

O panorama argentino revela um sistema avaliativo com tradição técnica, alta transparência e ampla divulgação pública. A avaliação Aprender apresenta maior estabilidade histórica e institucional, enquanto Aprender Alfabetização cumpre papel estratégico recente no âmbito da política de alfabetização.

As diferenças entre ambas se concentram sobretudo na maturidade institucional, na extensão da série histórica e na forma como os resultados são traduzidos para o debate público e incorporados ao ciclo de políticas.

**De modo geral, a Argentina conta com instrumentos que permitem leitura consistente do sistema educacional. O desafio reside em garantir a continuidade das avaliações, com comparabilidade no tempo — especialmente no caso da alfabetização — e avançar na integração mais sistemática entre avaliação e tomada de decisão.**

Como veremos nos próximos países da série, as avaliações em escala ocupam lugares distintos nos sistemas educacionais latino-americanos. Entender essas diferenças é fundamental para compreender não apenas os resultados, mas também a capacidade de cada país de transformar dados em política pública consistente.



## PARTE II – SISTEMAS NACIONAIS EM PERSPECTIVA

**Capítulo 5****Peru: institucionalização com tensões metodológicas**

O Peru consolidou, ao longo das últimas décadas, um sistema de avaliação educacional que se tornou um pilar importante para o monitoramento de políticas públicas. Com uma trajetória que remonta a 1996, o país consolidou a Evaluación Censal de Estudiantes (ECE) como principal instrumento nacional de monitoramento da aprendizagem, com aplicações regulares entre 2007 e 2019. A partir de 2023, o Peru passou a organizar suas medições sob um marco mais amplo, a Evaluación Nacional de Logros de Aprendizaje (ENLA), que articula diferentes modalidades de avaliação — censitárias e amostrais — e amplia o escopo analítico do sistema.

Nesse novo arranjo, a ENLA não substitui necessariamente a ECE, mas funciona como um guarda-chuva avaliativo, incorporando aplicações censitárias quando pertinente e ampliando a análise para fatores de contexto — incluindo dimensões socioemocionais — que influenciam o cotidiano escolar.

**Embora o sistema apresente um alto grau de institucionalização e continuidade entre diferentes governos, ele enfrenta desafios estruturais e metodológicos que impactam a leitura precisa dos avanços educacionais no longo prazo.**



Essa reorganização aumenta a flexibilidade do sistema, mas também exige cautela na leitura de tendências de longo prazo, uma vez que mudanças de desenho e cobertura podem afetar a comparabilidade direta com séries anteriores.

## Qualidade técnica e os desafios da comparabilidade

A qualidade técnica dos instrumentos peruanos é reconhecida, apresentando um alto grau de consenso nacional. A metodologia é considerada estável e está alinhada ao Currículo Nacional da Educação Básica. No entanto, especialistas apontam pontos de atenção que podem afetar a comparabilidade no espaço e no tempo — especialmente no contexto da reorganização recente do sistema avaliativo — fator que consideramos fundamental.

- **Alteração na definição de ruralidade:**

Em 2021, o Ministério da Educação (Minedu) modificou os critérios do que é considerado zona rural. Essa mudança incluiu um número significativamente maior de escolas nesta categoria, o que pode comprometer a comparabilidade direta dos resultados atuais com as edições anteriores.

- **Representatividade territorial limitada:**

Nos anos em que a avaliação é realizada de forma amostral (EM), os dados não possuem representatividade para o nível de UGEL (Unidades de Gestão Educacional Local). Isso cria uma lacuna para os gestores da ponta, que só obtêm dados específicos para suas jurisdições nos ciclos censitários — modalidade historicamente associada à ECE e que continua sendo utilizada dentro do marco ENLA.

- **Independência institucional:** A UMC (Oficina de Medición de la Calidad), responsável pelas medições, está vinculada diretamente ao Ministério. Apesar da sua tradição técnica, essa estrutura não oferece uma blindagem completa contra possíveis mudanças de prioridades associadas aos ciclos políticos.





## **Inclusão e Diversidade: O desafio da EIB**

Um diferencial relevante no processo avaliativo do Peru é o olhar para a Educação Intercultural Bilíngue (EIB). Desde 2007, o país realiza avaliações de leitura em línguas originárias para alunos do 4º ano, abrangendo idiomas como Quechua (em diversas variantes), Aimara, Awajún e Shipibo-Konibo.

Essa medição é fundamental em um país com tamanha diversidade cultural, mas impõe desafios logísticos e técnicos complexos. A avaliação busca garantir que o sistema não ignore as particularidades linguísticas de uma parcela significativa da população, embora a escala dessas aplicações e a produção de materiais específicos ainda enfrentem barreiras de implementação em larga escala.

## **A trajetória dos anos avaliados e o uso da tecnologia**

A série histórica peruana revela uma transição de foco ao longo dos anos. Entre 2007 e 2014, o sistema concentrou-se quase exclusivamente no 2º ano do Ensino Primário. Foi apenas em 2016 que o 4º ano passou a ser avaliado de forma censitária para o ensino regular, tornando-se, desde então, a série de referência para o monitoramento contínuo. Considerando o momento em que vários países da América Latina vem dando mais atenção à alfabetização inicial, inclusive o Peru, não ter avaliações com periodicidade e comparabilidade no 2º ano, certamente é um problema.

No que diz respeito à tecnologia, o Peru demonstra um estágio intermediário de transição. Durante os anos de pandemia, o país implementou a EVA (Evaluación Virtual de Aprendizajes). Embora tenha sido uma resposta necessária ao fechamento das escolas, o uso de plataformas digitais ainda é visto com cautela para garantir que a falta de infraestrutura tecnológica em áreas remotas não aprofunde as desigualdades já existentes nos resultados.

## Transparência e a "Educação em Dados"

O sistema peruano dedica esforços à democratização do acesso à informação. Os dados são traduzidos em escalas de níveis de conquista — como "Satisfatório", "Em processo" e "Em início" — que buscam tornar o dado estatístico mais inteligível para as famílias. O Minedu utiliza recursos diversificados, como podcasts e webinários, para ampliar o alcance do debate. Além disso, existe um portal de consulta dinâmica que permite o acesso a microdados históricos, integrando resultados nacionais e internacionais.



## A incidência dos resultados nas políticas públicas

O nível de incidência das avaliações no desenho de políticas públicas é considerado médio. No passado, os resultados fundamentaram decisões como o Bono de Incentivo al Desempeño Escolar (BDE), que bonificava docentes e diretores de escolas com avanços significativos. Os dados também orientavam o Programa Estratégico Logros de Aprendizaje (PELA), que direcionou apoio pedagógico e recursos orçamentários para as regiões com maiores lacunas de aprendizagem. Mais recentemente, o país passou a difundir "Kits de Avaliação Diagnóstica", permitindo que os professores utilizem os parâmetros da ENLA para monitorar o progresso dos alunos diretamente em sala de aula. Atualmente, a clareza sobre o uso efetivo das avaliações no desenho e acompanhamento das políticas educativas se dissipou. Essa situação constitui um verdadeiro desafio e uma oportunidade para o país.

## PARTE II – SISTEMAS NACIONAIS EM PERSPECTIVA

**Capítulo 6****Chile: maturidade, crise e reinvenção**

Se há um país na região que exemplifica o amadurecimento institucional de um sistema de avaliação — passando por fases de aceitação técnica absoluta, crises de legitimidade e, finalmente, uma sofisticada modernização tecnológica —, esse país é o Chile.



O sistema chileno de avaliação em larga escala articula-se em torno de um instrumento fundamental: o SIMCE (Sistema de Medición de la Calidad de la Educación). Diferente de outros vizinhos que ainda lutam para consolidar séries históricas, o Chile ostenta uma tradição que remonta a 1988. No entanto, o SIMCE não é apenas um sobrevivente do tempo; ele é um sobrevivente das transformações políticas e sociais do país. Sua trajetória é marcada por um deslocamento profundo: de um modelo de alta pressão, que nos anos 90 era visto com grande confiança pela sua precisão técnica, para um cenário contemporâneo de intenso debate sobre o "para quê" e o "como" avaliamos.

Essa evolução foi impulsionada por alguns questionamentos. No passado, o uso excessivo dos resultados para fins de ranqueamento e punição administrativa gerou efeitos colaterais severos, como por exemplo o empobrecimento do currículo escolar, já que muitas instituições, pressionadas pelos números, passaram a "ensinar para o teste", o que na prática levou à redução de tempo em algumas disciplinas para priorizar Matemática e Linguagem. Embora a rigidez do currículo atual limite essa prática, a crítica sobre o sacrifício curricular é uma parte importante da história do debate avaliativo daquele período.

Para salvar a legitimidade do instrumento e evitar que ele se tornasse um mecanismo de segregação, o Chile promoveu uma reforma essencial. Hoje, o sistema busca um maior equilíbrio. A avaliação não se restringe apenas ao SIMCE, integrando-se a um sistema mais amplo que inclui outras ferramentas, reduz o número de etapas avaliadas e incorpora os chamados Indicadores de Desenvolvimento Pessoal e Social, sinalizando que a qualidade educativa não se resume apenas à proficiência acadêmica. Também foi disponibilizada às escolas uma ferramenta de avaliação diagnóstica digital que devolve resultados imediatos aos docentes (DIA, Diagnóstico Integral de Aprendizagem).



**Um dos pontos de maior destaque no caso chileno, e que serve de referência para toda a América Latina, é o seu pioneirismo tecnológico.**

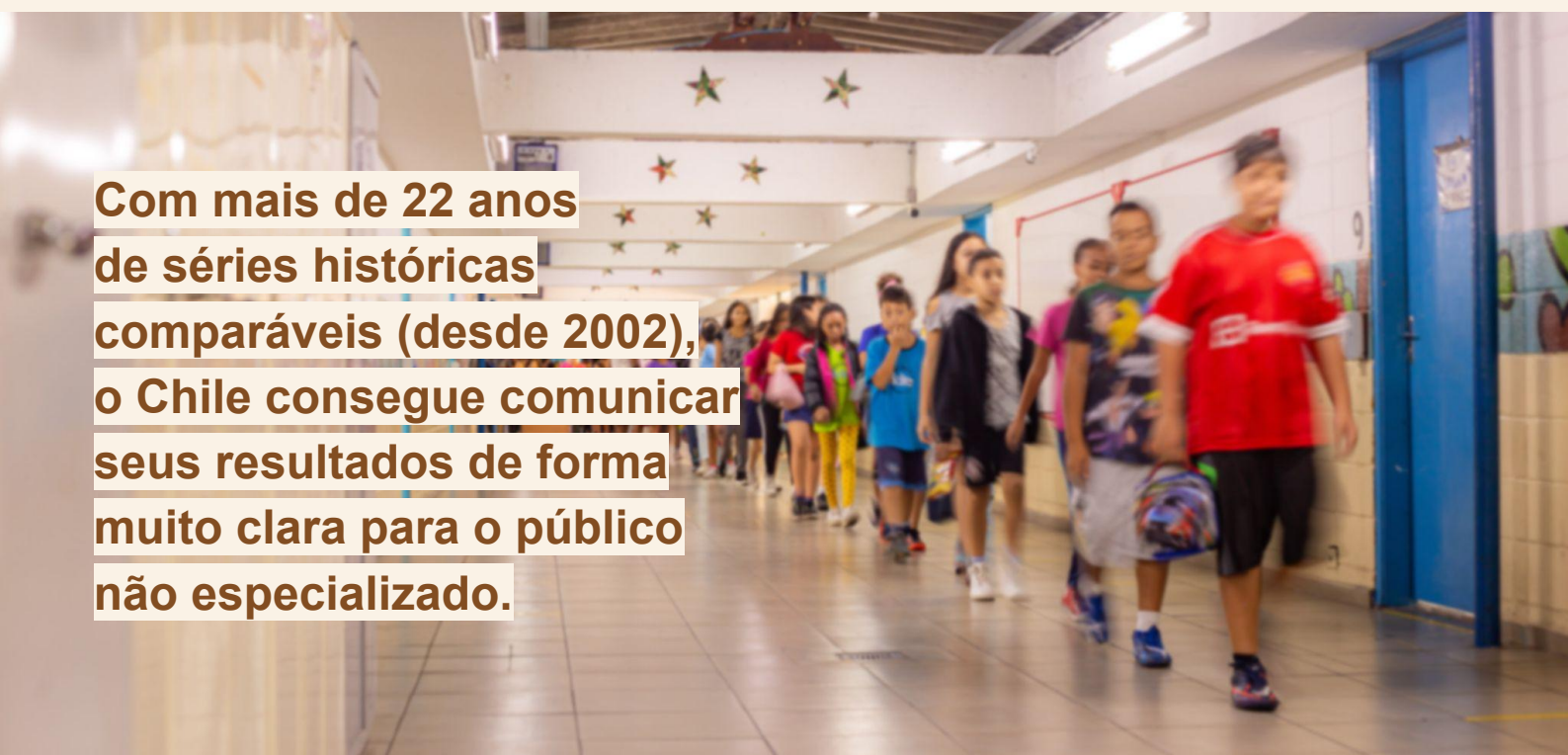
O Chile opera na vanguarda digital da avaliação, com um nível de sofisticação que permite reduzir o ciclo de processamento de dados de 220 para apenas 120 dias. Essa agilidade não é apenas burocrática; ela garante que o dado chegue à escola ainda com frescor pedagógico. Assim, por exemplo, a poucos dias de iniciado o ano letivo, já foram publicados os resultados do SIMCE 2025. Esta entrega rápida já acontece há quatro anos seguidos.

O uso de algoritmos de inteligência artificial para detectar anomalias e a criação de portais de resultados extremamente intuitivos e amigáveis para pais e diretores transformam o SIMCE em uma ferramenta de gestão viva. Além disso, o país deu passos largos na inclusão, utilizando softwares de adaptação e leitores de tela para que estudantes com deficiência possam ser avaliados com equidade.

Em termos de incidência nas políticas públicas, o SIMCE é o alicerce do sistema de accountability chileno. Se na Argentina o uso dos resultados ainda busca uma maior sistematização nas decisões políticas, no Chile ele é estruturante. Contudo, é necessário cautela para não sobredimensionar seu papel, que, embora importante, faz parte de um conjunto maior de evidências e ações políticas. Os dados alimentam diretamente a Agência de Qualidade da Educação, um órgão que goza de significativa autonomia e que utiliza as evidências para classificar e apoiar as escolas.



A institucionalização do sistema chileno também merece destaque. A avaliação é um pilar da Lei N° 20.529 e é executada pela Agência de Qualidade da Educação, um órgão que goza de autonomia e estabilidade técnica, protegendo o sistema de variações políticas bruscas. Ao contrário do que observamos em alguns contextos onde a avaliação é vista como uma ferramenta punitiva, o Chile tem buscado transformar o SIMCE em um mecanismo de apoio. O sistema classifica as escolas e direciona recursos e suporte técnico para aquelas em situação crítica, tornando a incidência em políticas públicas extremamente alta e estruturante.



**Com mais de 22 anos de séries históricas comparáveis (desde 2002), o Chile consegue comunicar seus resultados de forma muito clara para o público não especializado.**

As métricas de proficiência são traduzidas em níveis intuitivos — Insuficiente, Elemental e Adequado —, o que permite um controle social mais efetivo por parte da comunidade escolar. Ao olharmos para o Chile, vemos um sistema veterano que soube reconhecer seus limites e se modernizar, provando que a tecnologia e a transparência são os melhores antídotos para o desgaste de um sistema de avaliação em larga escala.

**Em conclusão, o caso chileno se erige como um referente de maturidade institucional e resiliência na América Latina. Apesar de sua complexa trajetória, o sistema de avaliação chileno demonstrou um compromisso com a transparência, a institucionalização robusta e um notável pioneirismo tecnológico, pilares que garantem dados confiáveis, ciclos de processamento ágeis e uma comunicação clara de resultados ao público não especializado.**

Estes alicerces transformam o SIMCE em uma ferramenta de gestão viva e de alto impacto nas políticas públicas. No entanto, os resultados recentes, como a estagnação em Leitura e a alta porcentagem de estudantes no nível 'Insuficiente' em áreas-chave, nos lembram que a persistência do desafio estrutural é inegável. A grande lição do Chile é que, embora a tecnologia e a solidez institucional sejam antídotos contra o desgaste do sistema, o sucesso final será medido pela sua capacidade de traduzir a evidência em uma melhoria pedagógica efetiva que impulse a qualidade e a equidade para todos os estudantes.

## PARTE II – SISTEMAS NACIONAIS EM PERSPECTIVA

**Capítulo 7****Colômbia: sistema estruturado, consistência e o desafio da última milha**

A Colômbia conta atualmente com um sistema consolidado de avaliação nacional de aprendizagem em escala, conhecido como Saber, que mede os resultados educacionais em diferentes etapas da trajetória escolar.

**Saber: um sistema estruturado de avaliação do sistema educacional**

As avaliações Saber constituem o principal instrumento nacional de medição de aprendizagem na Colômbia. Elas são aplicadas em diferentes momentos da trajetória escolar e fazem parte de um sistema integrado de avaliação educacional conduzido pelo ICFES (Instituto Colombiano para la Evaluación de la Educación).

No caso da educação básica, a avaliação ocorre principalmente nos 3º, 5º e 9º anos, permitindo acompanhar o desenvolvimento das aprendizagens ao longo do ensino fundamental. Essa estrutura cria um sistema relativamente abrangente de monitoramento do desempenho escolar.



Do ponto de vista técnico, as avaliações Saber possuem alto grau de reconhecimento entre especialistas, gestores e pesquisadores, com metodologia consolidada e continuidade ao longo do tempo. As críticas existentes concentram-se mais no uso pedagógico dos resultados e na periodicidade da aplicação.

Outro aspecto relevante é a ampla comparabilidade territorial. Os resultados são divulgados com desagregações por departamentos, municípios e tipos de escola, o que permite análises detalhadas das desigualdades educacionais dentro do país — uma característica especialmente importante em sistemas educacionais descentralizados.



## **Comparabilidade histórica e transparência dos dados**

Um dos pontos fortes do sistema colombiano é a continuidade das medições ao longo do tempo, o que permite a construção de séries históricas e a identificação de tendências no desempenho educacional.

A estabilidade metodológica e a regularidade das aplicações permitem que pesquisadores, gestores públicos e organizações da sociedade civil acompanhem a evolução dos resultados com relativa consistência.

A transparência dos dados também é elevada. O sistema disponibiliza relatórios técnicos, documentação metodológica e bases de dados que permitem análises mais aprofundadas. A divulgação pública dos resultados costuma ocorrer de forma ampla e regular, com repercussão no debate educacional e na mídia.

Outro elemento relevante é a forma de comunicação dos resultados. Embora as avaliações utilizem escalas técnicas de proficiência, o sistema também traduz os resultados em níveis de desempenho que facilitam a interpretação para gestores e para o público mais amplo.



## **Institucionalização e governança do sistema de avaliação**

O sistema Saber apresenta alto grau de institucionalização, com continuidade ao longo de diferentes governos e papel consolidado na arquitetura das políticas educacionais colombianas.

A responsabilidade técnica pela avaliação é exercida pelo ICFES, uma instituição com tradição técnica no país. Ainda que esteja vinculada ao governo, a organização construiu ao longo do tempo uma reputação de capacidade técnica e estabilidade institucional.

**Essa combinação entre tradição técnica, continuidade das medições e divulgação pública dos resultados contribui para que o sistema de avaliação desempenhe um papel estruturante na produção de informação sobre a educação colombiana.**

## Uso dos resultados e relação com políticas públicas

Os resultados das avaliações Saber são utilizados para diversas finalidades dentro do sistema educacional colombiano.

Entre os usos mais relevantes estão:

- o monitoramento do desempenho educacional em diferentes níveis do sistema;
- a identificação de desigualdades territoriais de aprendizagem;
- o apoio ao desenho e ajuste de políticas educacionais.

Esse uso relativamente estruturado da informação reforça o papel da avaliação como instrumento de gestão educacional e de accountability pública.

Ao mesmo tempo, como ocorre em muitos países, permanece o desafio de aprofundar a tradução dos resultados em ações pedagógicas concretas nas escolas e redes de ensino.





## Um sistema consolidado na região

O panorama colombiano revela um sistema de avaliação em escala relativamente consolidado na América Latina, com trajetória contínua, forte comparabilidade territorial e ampla disponibilidade de dados.

A avaliação Saber cumpre um papel central no acompanhamento do desempenho educacional do país, oferecendo informação relevante tanto para o debate público quanto para a formulação de políticas.

Os desafios futuros parecem estar menos ligados à construção do sistema avaliativo em si — que já apresenta bases técnicas sólidas — e mais relacionados à capacidade de transformar essas informações em melhorias sustentadas na aprendizagem dos estudantes.

Como veremos nos demais países dos textos desta mini-publicação, as avaliações em escala ocupam posições distintas dentro dos sistemas educacionais latino-americanos. Compreender essas diferenças ajuda a entender não apenas os resultados observados, mas também a capacidade de cada país de utilizar dados educacionais como instrumento efetivo de política pública.

## PARTE II – SISTEMAS NACIONAIS EM PERSPECTIVA

**Capítulo 8****México: a ausência como escolha e seus efeitos sobre o sistema**

Ao longo das últimas décadas, o México chegou a construir um dos sistemas de avaliação educacional mais estruturados da América Latina. Diferentes instrumentos nacionais — como EXCALE, ENLACE e posteriormente PLANEA — permitiram acompanhar a evolução dos resultados educacionais, produzir diagnósticos do sistema e alimentar o debate público sobre qualidade da educação.

**Atualmente, no entanto, o México não possui uma avaliação nacional padronizada de aprendizagem em larga escala com caráter somativo — isto é, uma avaliação capaz de medir de forma comparável o que os estudantes estão aprendendo ao longo do tempo e entre diferentes regiões do país.**



A trajetória mexicana em avaliação educacional começa ainda nos anos 1980, com a criação do Sistema Nacional de Evaluación. A partir dos anos 2000, o país avançou significativamente na institucionalização desse campo com a criação do Instituto Nacional para la Evaluación de la Educación (INEE), responsável por desenvolver avaliações nacionais e consolidar uma cultura de avaliação educacional.

**Nesse período, o país contou com avaliações nacionais relevantes, entre elas:**

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

Essas avaliações permitiam medir o desempenho dos estudantes em diferentes etapas da educação básica, com resultados comparáveis entre estados e ao longo do tempo. Paralelamente, o INEE se consolidou como um organismo técnico relativamente autônomo, responsável por garantir qualidade metodológica e transparência na divulgação dos dados.

Esse arranjo institucional aproximava o México de modelos observados em outros países da região que desenvolveram sistemas nacionais robustos de avaliação educacional.

## **Uma mudança de enfoque: da avaliação padronizada à avaliação diagnóstica**

Esse cenário começou a mudar a partir da reforma educacional de 2019. Neste momento, o INEE foi extinto e substituído pela Comisión Nacional para la Mejora Continua de la Educación (Mejoredu), que passou a ter um mandato mais restrito e menor autonomia institucional.

Posteriormente, em 2025, a própria Mejoredu também foi extinta no contexto de uma reforma mais ampla que eliminou diversos organismos autônomos ou descentralizados da administração pública. Como resultado, o país passou a não contar mais com uma instituição nacional dedicada à avaliação educacional com o mesmo desenho institucional anterior.



Esse processo acabou redesenhando o papel da avaliação educacional dentro do sistema mexicano.

Em paralelo a essas mudanças institucionais, também houve uma alteração importante na concepção de avaliação educacional adotada pelo país.

A política educacional associada ao modelo da Nueva Escuela Mexicana, lançada em 2022, passou a privilegiar um enfoque formativo e qualitativo da avaliação. Nesse modelo, o objetivo principal não é classificar ou comparar estudantes por meio de provas padronizadas, mas apoiar o processo de aprendizagem dentro da própria escola.

Nesse contexto, surgiram os Ejercicios Integradores del Aprendizaje (EIA), instrumentos diagnósticos aplicados no início do ciclo escolar. Esses exercícios são organizados por fases — que agrupam vários anos escolares — e buscam oferecer aos docentes informações sobre o nível de desenvolvimento dos estudantes em diferentes campos formativos.

### **O enfoque é essencialmente pedagógico:**

- **a aplicação é voluntária,**
- **os próprios docentes corrigem as atividades com base em rubricas,**
- **os resultados são utilizados principalmente dentro da escola para orientar o trabalho pedagógico.**



Esse modelo tem coerência com uma visão formativa da avaliação, cada vez mais presente em diversos sistemas educacionais.



## As limitações de modelo atual

Apesar de suas virtudes pedagógicas, o modelo vigente apresenta algumas limitações quando analisado sob a perspectiva do acompanhamento do sistema educacional.

**Primeiro**, os instrumentos atuais não foram concebidos para permitir comparações entre estados ou ao longo do tempo. O foco é o diagnóstico local — sala de aula, grupo ou escola — e não a medição do desempenho do sistema educacional como um todo.

**Segundo**, os resultados não são publicados em nível nacional ou subnacional, permanecendo principalmente como insumo para o trabalho pedagógico das escolas.

**Terceiro**, a responsabilidade pela avaliação passou a estar concentrada em áreas internas da Secretaria de Educación Pública, em um arranjo institucional diferente daquele que existia quando havia organismos especializados dedicados exclusivamente a essa função.

**Na prática, o país passou de um modelo fortemente estruturado em avaliações nacionais padronizadas para outro em que a avaliação tem um caráter mais pedagógico e situado no contexto das escolas.**

Não há dúvidas sobre a importância das avaliações formativas. Nesse sentido, o enfoque atualmente adotado no México possui muitos méritos. No entanto, avaliações de aprendizagem em escala, quando bem desenhadas e implementadas, cumprem outro papel igualmente fundamental. Elas são instrumentos imprescindíveis para orientar políticas públicas educacionais e acompanhar a evolução dos sistemas de ensino — especialmente em países da América Latina, cujos sistemas educacionais estão em um nível de maturidade institucional que torna esse tipo de monitoramento particularmente relevante.

## O impacto na produção de informação educacional

Essa mudança tem implicações importantes para a produção de informação sobre o sistema educacional.

Sem avaliações nacionais comparáveis, torna-se mais difícil responder a perguntas que normalmente orientam o desenho de políticas educacionais, como por exemplo:

- Como estão evoluindo as aprendizagens dos estudantes ao longo do tempo?
- Existem diferenças importantes entre estados ou regiões?
- Quais políticas educacionais estão gerando melhores resultados?

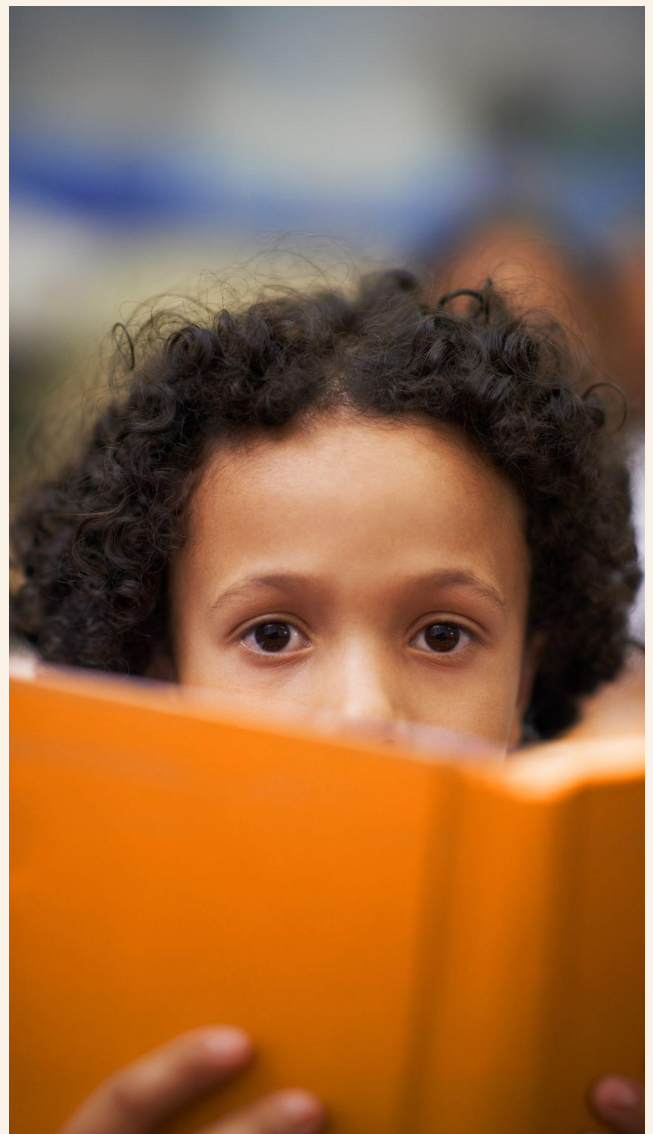
Essas perguntas continuam sendo relevantes para qualquer sistema educacional, e diferentes países têm buscado equilibrar o uso de avaliações formativas no nível da escola com instrumentos nacionais que permitam acompanhar tendências do sistema.

## O caso mexicano no contexto latino-americano

Quando comparado a outros países da região analisados nos capítulos anteriores — como Brasil, Peru ou Argentina — o México apresenta hoje uma situação particular.

Enquanto esses países mantêm avaliações nacionais periódicas capazes de produzir séries históricas e diagnósticos comparáveis, o México atravessa um período de redefinição do papel da avaliação educacional dentro de sua política pública.

Ao mesmo tempo, permanece um legado institucional e técnico relevante acumulado ao longo de décadas. Parte dessa informação histórica, inclusive, vem sendo preservada por organizações da sociedade civil que buscam evitar a perda desse patrimônio de dados educacionais.



## PARTE II – SISTEMAS NACIONAIS EM PERSPECTIVA

## Capítulo 9

## Um quadro comparado para a América Latina

Item de análise	Brasil		Argentina		Perú	Chile	Colômbia	México
Avaliação	SAEB Alfabetização	SAEB	APRENDER Alfabetização	APRENDER	ENLA	SIMCE	SABER	NÃO HÁ
1. Séries/anos avaliados	2º ano do EF	5º e 9º anos EF; 3º ano do EM	3º ano do primário	6º ano do primário	2º e 4º do primário; 2º do secundário	4º, 6º e 8º Básica	Educação Básica – Anos Iniciais: 3º e 5º anos	
2. Qualidade técnica (grau de consenso nacional)	Média	Média	Média	Média	Alta	Média	Alta	
3. Periodicidade	Anual	Bienal	Não definida	Bienal	Anual (alternados)	Anual (4º) / Bienal	A cada 2 ou 3 anos	
4. Abertura e comparabilidade por entidades subnacionais ou regionais	Alta	Alta	Média	Alta	Média	Alta	Alta	
5. Comparabilidade de tempo	Média	Alta	Baixa	Alta	Média	Alta	Alta	
6. Transparência de dados e metodologia	Alta	Alta	Alta	Alta	Alta	Alta	Média	
7. Independência técnica e institucional do organismo responsável pela medição	Alta	Alta	Média	Média	Média	Média/Alta	Alta	
8. Publicação de resultados	Alta	Alta	Alta	Alta	Alta	Alta	Média	
9. Grau de institucionalização	Médio	Alto	Baixo	Alto	Alto	Alto	Medio	
10. Uso coerente de tecnologia	Médio	Médio	Médio	Médio	Alto	Alto	Médio	
11. Anos de histórico	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 inteligíveis para públicos não especializados	Alto	Médio	Baixo	Médio	Alto	Alto	Médio	
13. Uso dos resultados e nível de incidência em políticas públicas	Alto	Alto	Alto	Médio	Médio	Alto	Médio	



## Critérios para avaliações subjetivas

### Qualidade técnica (grau de consenso nacional)

**Alto consenso:** Avaliação amplamente reconhecida por especialistas, academia, gestores e órgãos técnicos; metodologia estável; críticas concentradas em uso, não no instrumento.

**Consenso intermediário:** Reconhecimento técnico relevante, mas com controvérsias recorrentes sobre desenho, metodologia ou interpretação.

**Baixo consenso:** Questionamentos frequentes sobre validade, confiabilidade ou adequação técnica do instrumento.

### Abertura e comparabilidade por entidades subnacionais ou regionais

**Alta:** Resultados comparáveis entre estados/municípios/regiões, com recortes públicos e metodologicamente consistentes.

**Média:** Comparabilidade possível apenas para alguns níveis ou com limitações relevantes.

**Baixa ou inexistente:** Resultados agregados nacionalmente, sem possibilidade real de comparação subnacional.

### Comparabilidade no tempo

**Alta:** Escala e metodologia estáveis; séries históricas contínuas; mudanças raras e bem documentadas.

**Média:** Comparabilidade possível, mas com rupturas metodológicas pontuais.

**Baixa:** Mudanças frequentes que dificultam a leitura de tendência.

## Transparência de dados e metodologia

**Alta:** Microdados, relatórios técnicos e documentação metodológica públicos e acessíveis.

**Média:** Relatórios disponíveis, mas com limitações de acesso a dados ou detalhes metodológicos.

**Baixa:** Pouca informação pública sobre metodologia ou dados.

## Independência técnica e institucional do órgão responsável pela medição

**Alta:** Órgão técnico com autonomia institucional, regras claras de governança e proteção contra interferência político-partidária.

**Média:** Órgão ligado ao governo, mas com tradição técnica consolidada e algum grau de blindagem.

**Baixa:** Forte dependência política, com mudanças frequentes associadas a ciclos de governo.

## Publicização dos resultados

**Alta:** Divulgação ampla, regular e acessível, com relatórios públicos e comunicação ativa.

**Média:** Divulgação pública, mas com alcance restrito ou pouca sistematicidade.

**Baixa:** Divulgação limitada ou restrita a públicos técnicos.

## Grau de institucionalização

**Alta:** Avaliação prevista em lei ou normativa estável; continuidade entre governos.

**Média:** Avaliação recorrente, mas baseada em decisões administrativas.

**Baixa:** Avaliação pontual ou sujeita a descontinuidade política



## Uso coerente de tecnologia

### Alto:

- A tecnologia é usada de forma alinhada à finalidade da avaliação (monitoramento, comparabilidade, política pública).
- Contribui claramente para: redução de custos; maior escala ou cobertura; ciclos avaliativos mais curtos; melhor comunicação dos resultados.
- Não compromete: comparabilidade no tempo; equidade de acesso; validade do instrumento.
- A tecnologia é subordinada ao desenho pedagógico e técnico, não o contrário.

### Médio:

- A tecnologia é utilizada, mas com ganhos limitados ou concentrados em apenas uma dimensão (ex.: logística); ou ainda em fase de transição/piloto.
- Há benefícios claros, mas também: restrições de acesso; limitações técnicas; ou dependência de infraestrutura desigual.
- O uso tecnológico não distorce o objetivo, mas ainda não o potencializa plenamente.

### Baixo:

- Uso tecnológico: inexistente; ou desarticulado da finalidade da avaliação; ou adotado mais por modismo do que por necessidade técnica.
- Pode gerar: aumento de desigualdades; perda de comparabilidade; fragilidade técnica ou operacional.
- A tecnologia se sobrepõe ao desenho avaliativo, em vez de servi-lo.

## Resultados inteligíveis para públicos não especializados

**Alta:** Resultados comunicados em métricas intuitivas (percentuais, níveis claros), compreensíveis para o público geral.

**Média:** Resultados compreensíveis com mediação técnica.

**Baixa:** Resultados apresentados apenas em escalas técnicas ou pontuações abstratas.

## Uso dos resultados e nível de incidência em políticas públicas

**Alto:** Resultados claramente utilizados para desenho, ajuste ou monitoramento de políticas públicas.

**Médio:** Uso parcial ou indireto, sem vínculo sistemático com decisões.

**Baixo:** Resultados pouco utilizados ou meramente informativos.



## Conclusão

# O futuro das avaliações em larga escala

Ao longo desta mini-publicação, procuramos discutir o papel das avaliações em larga escala na América Latina a partir de uma perspectiva menos polarizada e mais ancorada na realidade dos sistemas educacionais. Esses textos foram construídos em diálogo com Karina Stocovaz, Superintendente do Instituto Natura para os países de língua hispânica, cuja contribuição foi fundamental para ampliar o olhar regional.

As avaliações, como vimos, estão longe de serem instrumentos perfeitos. Elas não capturam toda a complexidade da educação, podem induzir distorções quando mal utilizadas e não substituem políticas públicas bem desenhadas. Ao mesmo tempo, sua ausência compromete dimensões centrais do funcionamento dos sistemas educacionais: a capacidade de diagnóstico, a transparência e o próprio sentido de responsabilização pública.



**A experiência latino-americana revela um cenário diverso, com países em diferentes estágios de desenvolvimento de seus sistemas avaliativos. Essa heterogeneidade não deve ser vista apenas como um desafio, mas também como uma oportunidade.**

Há espaço concreto para aprendizado entre países, circulação de boas práticas e aceleração de trajetórias a partir de experiências já consolidadas na região.

**Avançar, nesse contexto, não significa simplesmente expandir avaliações, mas aprimorar continuamente seu desenho, seu uso e sua integração com as políticas educacionais. Significa também reconhecer seus limites, evitando tanto a supervalorização quanto o descarte precipitado.**

Se bem construídas e utilizadas com responsabilidade, as avaliações em larga escala seguem sendo um dos poucos instrumentos capazes de sustentar, ao mesmo tempo, visão sistêmica, coordenação e pressão por melhoria. Não resolvem, por si só, os desafios educacionais da região — mas são parte indispensável de qualquer estratégia que pretenda enfrentá-los com seriedade e consistência.



instituto  
natura



AVON