



Organização  
das Nações Unidas  
para a Educação,  
a Ciência e a Cultura



Instituto Internacional de  
Planejamento Educacional  
IIPE UNESCO Buenos Aires  
Escritório para a América Latina



ESTADO DA ARTE  
DA PESQUISA  
À POLÍTICA

# Uso de resultados de provas de aprendizagem na formulação de políticas educacionais na América Latina

---

Cecilia Galas Taboada,  
María Teresa Gutiérrez Cortés  
e Eva María Hamilton Vélez

Março 2020

### Advertência

O uso de linguagem não discriminatória em função do gênero das pessoas é uma das prioridades do Escritório para a América Latina do Instituto Internacional de Planejamento Educacional da UNESCO. No entanto, ainda não existe um consenso em relação às diferentes soluções propostas para sua aplicação na língua portuguesa.

Nesse sentido, evitamos, nos nossos textos, o uso de expressões tradicionais que já foram abandonadas no português acadêmico e profissional contemporâneo, como por exemplo a palavra “homem” para designar “seres humanos”. Além disso, sempre que possível, buscamos empregar palavras de gênero neutro como estratégia de substituição de palavras em masculino.

Porém, com o propósito de evitar sobrecarga gráfica a partir da duplicação lexical em artigos, substantivos e adjetivos para destacar a existência de diferentes gêneros, optamos pelo uso genérico do masculino, entendendo que todas as suas ocorrências representam, sempre, todas as pessoas.

Os termos empregados e os dados apresentados ao longo desta publicação não implicam qualquer posicionamento por parte da UNESCO ou do IIPE, no que diz respeito ao estatuto jurídico de países, territórios, cidades, regiões ou suas autoridades, fronteiras ou limites.

Publicado em 2020 pelo Escritório para a América Latina do Instituto Internacional de Planejamento Educacional da Organização das Nações Unidas para a Educação, a Ciência e a Cultura.

Agüero 2071, Cidade Autônoma de Buenos Aires, Argentina.

© UNESCO 2020

-----

Esta publicação está disponível em acesso livre sob a licença Attribution-ShareAlike 3.0 IGO

(CC BY-SA 3.0 IGO) (<http://creativecommons.org/licenses/by-sa/3.0/igo/>). Ao utilizar o conteúdo desta publicação, os usuários aceitam as condições de uso do Repositório UNESCO sobre acesso livre (<http://www.unesco.org/open-access/terms-use-ccbysa-sp>).



A série de documentos *Estado da Arte de pesquisas sobre políticas educacionais* faz parte das atividades da área de Pesquisa e Desenvolvimento do Escritório para a América Latina do Instituto Internacional de Planejamento Educacional (IIPE), cuja missão é fortalecer capacidades dos Estados-membros da UNESCO na região para planejar e gerir seus sistemas educacionais.

A produção desses documentos se articula especificamente com o projeto Sistema de Informação de Tendências Educacionais na América Latina (SITEAL). O SITEAL é um observatório online que oferece uma base de políticas, regulamentos, pesquisas e estatísticas sobre a situação educacional na região da América Latina, produzindo documentos que sistematizam e interpretam essas informações para seu acompanhamento.

O objetivo dos Estados da arte é fornecer às equipes técnicas do governo, pesquisadores, professores e capacitadores, documentos que sistematizem e analisem pesquisas sobre temas relevantes para o planejamento e a gestão das políticas educacionais. Através desse tipo de estudo, há um esforço em produzir conhecimento útil a fim de identificar e resumir o nível de desenvolvimento de cada tema, reconhecendo dimensões, categorias e focos, bem como pontos de dissidência, as constantes, as tendências e os desafios atuais que surgem em cada caso. O objetivo desses estudos é contribuir para a compreensão do tema selecionado e, ao mesmo tempo, enriquecer o debate educacional.

Com isso, espera-se que essa série de documentos, através de seus diferentes eixos, funcione como um conjunto de estudos que facilite a tradução do conhecimento produzido pelo mundo acadêmico para o campo das políticas públicas.

## Conteúdo

1. Introdução .....	5
1.1 Objetivos e perguntas orientadoras para a revisão da literatura .....	5
1.2 Escopo e metodologia da revisão .....	6
1.3 Estrutura da revisão .....	7
2. Visão geral do desenvolvimento de sistemas de avaliação na região .....	8
2.1 A criação dos sistemas de avaliação: contexto e motivações .....	9
2.2 Primeira onda: surgimento de avaliações em larga escala e criação de sistemas de avaliação .....	15
2.3 Segunda onda: evolução institucional, consideração de fatores associados e feedback de resultados por níveis de realização .....	16
2.4 Terceira onda: desenvolvimento de medições mais complexas, incorporação de elementos não cognitivos na avaliação e preocupação pelo uso dos resultados .....	17
2.5 Marcos institucionais das unidades de avaliação .....	19
3. Marco conceitual sobre os usos dos resultados das avaliações .....	23
3.1 Tipologia dos usos da evidência .....	23
3.2 Objetivos da avaliação .....	24
3.3 Interação entre visões e funções da avaliação e o uso dos resultados .....	25
3.4 Teorias da mudança das avaliações em larga escala .....	27
4. O uso dos resultados da avaliação em políticas e intervenções educacionais na região .....	29
4.1 Definição da agenda política .....	30
4.2 Planejamento estratégico .....	31
4.3 Direcionamento de programas e intervenções .....	31
4.4 Orientação do currículo e materiais .....	32
4.5 Incentivos docentes .....	32
4.6 Classificação ou classificação das escolas .....	32

5. Fatores que influenciam o uso (ou o não uso) dos resultados das avaliações .....	33
5.1 Fatores políticos .....	34
5.2 Elementos institucionais .....	37
5.3 Aspectos técnicos.....	39
5.4 Comunicação e divulgação.....	42
6. Posicionamentos quanto às avaliações padronizadas .....	44
6.1 Teórico-ideológica .....	44
6.2 Técnico-instrumental.....	47
7. Conclusões.....	52
8. Bibliografia.....	

# 1. Introdução

Ao longo da década de 1990, houve grandes esforços para construir sistemas de avaliação em larga escala na América Latina. Quase trinta anos após o início de sua implementação, é importante examinar quais foram os avanços, aprendizados e desafios para definir a agenda política e acadêmica da década seguinte. Nos últimos vinte anos, foram realizados estudos que se aprofundaram nessa trajetória de desenvolvimento e consolidação de sistemas de avaliação na região. Destacam-se os esforços do Programa de Promoção da Reforma Educacional da América Latina e do Caribe (PREAL) durante os anos 2000 e, mais recentemente, do Banco Mundial e do Centro de Implementação de Políticas Públicas para Equidade e Crescimento (CIPPEC). Esta revisão da literatura busca contribuir com a integração – e, quando pertinente, com a atualização – dos diferentes marcos da conversa sobre o uso dos resultados das avaliações na região para identificar os temas prioritários no momento.

## 1.1. Objetivos e perguntas orientadoras para a revisão da literatura

O objetivo desta revisão da literatura é explorar o estado da arte sobre o uso dos resultados das avaliações de aprendizagem como instrumentos de melhoria nos países da América Latina. O foco deste documento consiste em identificar padrões em uso, elementos que influenciam os usos observados, bem como as diferentes posições que existem atualmente em torno das avaliações em larga escala.

As sínteses e conclusões apresentadas servirão para identificar tópicos específicos, nos quais será pertinente um maior aprofundamento para reconhecer fatores e dinâmicas que favorecem ou inibem o uso desses resultados para uma melhoria, considerando a complexidade do processo educacional e da elaboração de políticas, a própria história da região, bem como os diferentes atores envolvidos.

Objetivamente, as perguntas a serem respondidas são as seguintes:

- Como os sistemas de avaliação se desenvolveram na América Latina e quais foram os principais fatores que levaram à sua evolução?

- Quais são as referências de modelos de uso de resultados e que foram utilizados pelos países da região?
- Como os resultados da avaliação influenciaram políticas e intervenções nos países da região?
- Que fatores influenciam o uso (ou o não uso) dos resultados?
- Que posições existem na região atualmente em comparação com os sistemas de avaliação padronizados?

## 1.2 Escopo e metodologia da revisão

A seleção dos documentos para a revisão foi realizada de acordo com os seguintes critérios:

- Trabalhos publicados nos últimos quinze anos.
- O conteúdo aborda explicitamente a questão do uso dos resultados das avaliações. Nesse sentido, foram descartadas descrições detalhadas dos sistemas de avaliação, discussões técnicas sobre as características das provas e, de forma geral, relatórios descritivos dos resultados. Foram feitas exceções a estes últimos quando alguma seção do relatório sugeria usos para a informação apresentada.
- Os casos referem-se, principalmente, a experiências em países latino-americanos. Embora sejam incluídos alguns documentos do Programa de Avaliação Internacional de Estudantes (PISA) e de referências internacionais no uso das avaliações, foram descartados trabalhos que abordas sem o tema do uso em perspectiva internacional. Desta forma, favorecemos uma visão mais focada nos acertos e desafios particulares da região.
- Foram incluídos trabalhos de pesquisa tanto teórica quanto empírica.
- De acordo com as perguntas orientadoras, vários tipos de texto foram incluídos na revisão: publicações acadêmicas, artigos de opinião, relatórios institucionais, capítulos de livros e apresentações em seminários ou conferências.

É de particular interesse representar neste trabalho as diferentes vozes que participam da conversa regional sobre a relevância e o uso das avaliações em larga escala. Por esse motivo, em alguns casos excepcionais,

foram incluídos artigos que, sem atender aos padrões convencionais de rigor técnico, apresentam posições relevantes para essa conversa.

A extensão da literatura revisada foi delimitada de acordo com a observação de redundâncias nas referências e com a saturação de argumentação de um tema. Ainda assim, como é natural em trabalhos desta natureza, obras relevantes podem ter sido excluídas, e que deverão ser recuperadas em estudos posteriores.

### 1.3 Estrutura da revisão

A revisão se estrutura de maneira temática, integrando as contribuições e perspectivas de diferentes autores sobre um tema específico, de acordo com seu posicionamento.

A primeira seção sintetiza a evolução dos sistemas de avaliação na região para apresentar as descobertas explicitadas no restante do documento a respeito do uso dos resultados, das suas consequências nos sistemas educacionais e da própria avaliação, bem como críticas e recomendações feitas sob distintas perspectivas.

A segunda seção apresenta uma série de referências conceituais que embasaram a discussão sobre o uso dos resultados na região. Os diferentes marcos tratam tanto do uso de evidências e informações para a formulação de políticas quanto das diferentes lógicas de mudança subjacentes aos sistemas de avaliação.

A terceira seção é dedicada à descrição dos usos dos resultados de avaliações em larga escala observados na região, apresentando exemplos e casos de países com diferentes trajetórias em avaliação.

Na quarta seção, são abordados fatores que influenciam a forma de uso dos resultados, novamente, a partir de práticas documentadas em países da América Latina.

A quinta seção reúne as posições de vários autores em relação à relevância, implementação, uso e consequências das avaliações em larga escala, identificando pontos de controvérsia e consenso.

Finalmente, as conclusões reúnem as principais ideias que emergem da revisão da literatura e propõem linhas de pesquisa futuras.



## 2. Visão geral do desenvolvimento de sistemas de avaliação na região

Ao finalizar as duas primeiras décadas do século XXI, aproximadamente 65 milhões de adolescentes e jovens frequentam a educação secundária na América Latina. Em 13 dos 19 países da região esse nível educacional foi considerado obrigatório de modo completo e, nos países restantes, com exceção da Nicarágua, a obrigatoriedade alcança seus primeiros anos (secundária inferior ou ciclo básico, conforme a denominação em cada país). Apesar do esforço de inclusão, 15% da população entre 11 e 17 anos ainda não frequenta um estabelecimento de ensino e uma parte significativa da população adulta não concluiu esse nível educacional (SITEAL, 2019).

Para entender adequadamente as características, abordagens e usos que foram e estão sendo dados atualmente à avaliação e seus resultados, é necessário resgatar o contexto histórico em que surgem os sistemas de avaliação na América Latina e os principais momentos de sua evolução. O reconhecimento da influência do contexto político, educacional e técnico no qual as primeiras avaliações foram elaboradas ajuda a explicar alguns dos acertos, desafios, apoios e controvérsias que foram gerados em torno do uso dos resultados de provas padronizadas.

No documento *“As avaliações educacionais que a América Latina precisa”* (Ravela, et al., 2007), os autores sintetizam com precisão o caminho dos sistemas de avaliação educacional na região desde a década de noventa até 2007, distinguindo o comportamento entre países: uns que conservam seus sistemas com algumas mudanças, outros com descontinuidade nas aplicações, e outros tantos que reiniciam e repensam as abordagens, muitas vezes por motivações políticas.

Em seu livro *América Latina depois do PISA. Lições aprendidas da educação em sete países* (2015), Rivas enquadra o desenvolvimento de sistemas de avaliação dentro de uma análise das tendências políticas educacionais na região.

Da mesma forma, Felipe Martínez Rizo e Guillermo Ferrer fazem um relato do surgimento das avaliações padronizadas nacionais nos países da América Latina, bem como de sua evolução ao longo do tempo (Martínez Rizo F., 2008; Martínez Rizo F., 2016; Ferrer, 2006; Ferrer & Fiszbein, 2015).

A seguir, descrevemos brevemente os principais marcos no desenvolvimento de sistemas de avaliação na região, e apresentamos algumas interpretações da estrutura político-histórica em que surgem.

## 2.1 A criação dos sistemas de avaliação: contexto e motivações

Embora o Brasil, a Costa Rica e o Chile tenham se envolvido com o tema das avaliações em larga escala a partir de meados da década de 1980, as avaliações padronizadas foram implementadas em países da América Latina com maior força na década de 1990 (Martínez Rizo, 2008).

Isso envolveu a criação ou o fortalecimento de equipes técnicas especializadas em avaliação e o surgimento ou ampliação de funções das instâncias de avaliação. Em vários casos, a criação do sistema de avaliação como tal foi posterior ao início da aplicação das provas em larga escala.

**Tabela 1.** Início da aplicação de avaliações padronizadas na região sobre a avaliação da qualidade educacional

Ano da primeira avaliação em larga escala através de um sistema de avaliação da qualidade educacional <sup>1</sup>	Países
1982	Chile <sup>2</sup>
1986	Costa Rica
1990	Brasil
1991	Colômbia, República Dominicana
1992	Guatemala
1993	Argentina, El Salvador
1996	Cuba, Equador, Nicarágua, Paraguai, Peru, Uruguai
1997	Bolívia, Honduras
1998	México, Venezuela <sup>3</sup>
2000	Panamá

Elaboração própria com base na obra "A avaliação da aprendizagem na América Latina" (Martínez Rizo, 2016).

<sup>1</sup> Desde a década de 1960, vários países aplicaram provas de larga escala ou padronizadas para diferentes fins. No entanto, elas não foram adequadamente enquadradas nos sistemas de avaliação com foco na medição da qualidade educacional.

<sup>2</sup> Em 1968, a primeira avaliação padronizada foi realizada no país e em 1982 foram aplicadas as primeiras provas do que seis anos mais tarde seria o Sistema de Medição da Qualidade da Educação (SIMCE).

<sup>3</sup> O levantamento de 1998 foi o único exercício de avaliação em larga escala realizado no país até o momento.

Um esforço importante na agenda de avaliação regional foi a criação do Laboratório Latino-Americano de Avaliação da Qualidade da Educação (LLECE) em 1994, com o triplo objetivo de promover políticas baseadas em evidências por meio da produção de informações sobre a qualidade da educação e fatores associados, desenvolver capacidades de avaliação educacional e servir como fórum para a troca de ideias e melhores práticas em educação (UNESCO, 2014). Em 1997, os países-membros do LLECE implementaram o Primeiro Estudo Regional Comparativo e Explicativo (PERCE) sobre o desempenho de aprendizagem em leitura e de matemática para alunos da terceira e quarta séries da primária. O PERCE foi seguido por mais três estudos em que os graus de aplicação foram alternados – a quarta série foi trocada pela sexta – e os campos de conhecimento avaliados foram ampliados – adicionou-se a prova de ciências para alunos da sexta série da primária (Jiménez Moreno, 2016; UNESCO, 2014).

**Tabela 2.** Participação dos países nas avaliações do LLECE

Estudo	Ano de aplicação	Número de países participantes	Lista de países
PERCE	1997	13	Argentina, Bolívia, Brasil, Chile, Colômbia, Costa Rica, Cuba, Honduras, México, Paraguai, Peru, República Dominicana e Venezuela
SERCE	2006	16	Argentina, Brasil, Chile, Colômbia, Costa Rica, Cuba, Equador, El Salvador, Guatemala, México, Nicarágua, Panamá, Paraguai, Peru, República Dominicana e Uruguai <sup>4</sup>
TERCE	2013	15	Argentina, Brasil, Chile, Colômbia, Costa Rica, Equador, Guatemala, Honduras, México, Nicarágua, Panamá, Paraguai, Peru, República Dominicana e Uruguai <sup>5</sup> e o estado mexicano de Nuevo León

<sup>4</sup> Bolívia, Honduras e Venezuela não participaram deste segundo estudo, enquanto Equador, El Salvador, Guatemala, Nicarágua, Panamá e Uruguai aderiram.

<sup>5</sup> Apesar de ter participado dos dois estudos anteriores, Cuba decidiu não realizar o terceiro estudo.

ERCE 2019	2019	18	Argentina, Brasil, Bolívia, Chile, Colômbia, Costa Rica, Cuba, Equador, Guatemala, Honduras, El Salvador, México, Nicarágua, Panamá, Paraguai, Peru, República Dominicana e Uruguai <sup>6</sup>
-----------	------	----	--

Elaboração própria a partir de dados do LLECE.

Com a criação do PISA, a Organização para a Cooperação e o Desenvolvimento Econômico (OCDE), em 2000, foram estabelecidas políticas para uma avaliação padronizada na região, e essa prova internacional se colocou no centro das discussões sobre a política educacional e outros setores, por ser uma ferramenta que permite fazer comparações regional e globalmente.

**Tabela 3.** Aumento da participação de países da região no PISA

Ano de aplicação	Número de países participantes	Número de países participantes na região	Lista de países da região
2000	32	5	Argentina, Brasil, Chile, México y Peru <sup>7</sup>
2003	38	3	Brasil, México y Uruguai
2006	51	6	Argentina, Brasil, Chile, Colômbia, México, Uruguai
2009	66	8	Argentina, Brasil, Chile, Colômbia, México, Panamá, Peru e Uruguai
2012	57	8	Argentina, Brasil, Chile, Colômbia, Costa Rica, México, Peru, Uruguai
2015	72	9	Argentina, Brasil, Chile, Colômbia, Costa Rica, República Dominicana, México, Peru e Uruguai <sup>8</sup>
2018	79	9	Argentina, Brasil, Chile, Costa Rica, República Dominicana, México, Panamá, Peru e Uruguai

El Salvador também não participou e Honduras se incorporou pela primeira vez.

<sup>6</sup> Cuba, El Salvador e Bolívia se reincorporaram ao estudo. É a primeira avaliação internacional da qual a Bolívia participa desde 1997.

<sup>7</sup> Chile e Peru implementaram a avaliação em 2002 e Argentina em 2003 (OCDE, 2002).

<sup>8</sup> A Venezuela continua sendo um país-membro da LLECE, no entanto, não participou da aplicação do ERCE 2019.

PISA-D (2018)	8	5	Equador, Guatemala, Honduras, Panamá e Paraguai
---------------	---	---	---

*Elaboração própria a partir dos relatórios do PISA*

Em seu artigo “Is it worth it? Rationales for (non) participation in international large-scale assessments” (2019), Camilla Addey e Sam Sellar retomam de Steiner-Khamsi o agrupamento de motivações para participar de avaliações internacionais em quatro razões ou dimensões, e as desenvolvem da seguinte forma:

- **Dimensão política:** concentra-se em fatores políticos nos âmbitos nacional, regional e internacional para mobilizar ideias e legitimar decisões e políticas. As avaliações são instrumentos para criar conjunturas e janelas de oportunidades políticas. A narrativa construída é mais valorizada do que inferências válidas a partir dos dados. Também está relacionado a discursos de prestação de contas e transparência entre atores: governo, grupos de interesse, sociedade, professores, etc.
- **Dimensão econômica:** baseia-se na relação entre a teoria do capital humano e a crença de que o aprimoramento das habilidades medidas nas avaliações internacionais produzirá crescimento econômico (Hanushek e Woessmann, 2010). Com base nessa premissa, os resultados de provas como o PISA tornam-se sinais para os investidores, que as interpretam como indicadores de fator humano capacitado. Em alguns casos, o condicionamento do apoio internacional à disponibilidade de avaliações é um motor para o desenvolvimento desses sistemas. Por fim<sup>9</sup>, e não menos importante, é o crescimento da indústria avaliativa que se desenvolveu em torno da aplicação dessas provas e que proporciona sua permanência e crescimento.
- **Dimensão técnica:** a participação nessas avaliações é percebida como uma oportunidade para o desenvolvimento de capacidades avaliativas ou, quando apropriado, para compensar a falta delas. Além disso, as reuniões e oficinas de preparação para as provas proporcionam às equipes técnicas e tomadores de decisão de diferentes países a oportunidade de discutir e trocar ideias sobre políticas educacionais, planejamento e pesquisa, além do fato de que, em alguns casos, estabeleceu-se um sistema

<sup>9</sup> Empresas e consultores que oferecem serviços especializados para as agências responsáveis pelo desenvolvimento e aplicação das provas padronizadas.

de tutela entre países com mais experiência e países que estão começando ou tentando consolidar seus sistemas de avaliação.

- Dimensão sociocultural: também conhecida como “modelos de emulação normativa”, a adoção de políticas educacionais, neste caso de avaliação, visa legitimar o Estado e apresentá-lo como moderno, responsável e comprometido com a comunidade internacional.

Na realidade, as motivações incorporadas nessas dimensões estão entrelaçadas e podem mudar com o tempo, uma vez que o processo de formulação de políticas, incluindo políticas de avaliação, é confuso, pouco estruturado e impulsionado por muitos fatores, não apenas o técnico (Addey e Sellar, 2019; Porter, 2010; Ravela, 2015).

O marco motivacional usado por Addey e Sellar também é válido para explicar a incursão de países em avaliações nacionais em larga escala, uma vez que as dinâmicas políticas, econômicas, técnicas e socioculturais que operam internacionalmente são refletidas no interior dos países<sup>10</sup>.

Antes da década de 1990, os esforços da maioria dos países da região estavam concentrados na cobertura do sistema educacional, entendida como a matrícula nos níveis obrigatórios<sup>11</sup>. Sánchez Cerón e Corte Cruz del Sagrario identificam três características no contexto regional que marcaram a década de 90: a adoção de políticas neoliberais para reverter as crises econômicas, a reestruturação do Estado e a abertura de mercados, e uma diminuição nos gastos sociais favorecendo o investimento privado (Sánchez Cerón e Corte Cruz del Sagrario, 2013). No mesmo artigo, os autores denunciam a desigualdade social que se acentuou nas últimas décadas.

Jiménez Moreno (2016) argumenta que foi a abertura dos governos regionais ao mercado global, na mesma década, que despertou o interesse pela competitividade e focou a atenção na qualidade dos sistemas educacionais, com o subsequente investimento nos sistemas de avaliação para tentar medi-lo.

<sup>10</sup> Por exemplo, tanto no México quanto no Chile os resultados da avaliação foram chave para gerar o sentido de urgência necessária para aprovar as reformas educacionais estruturais de 2011 e 2013, respectivamente (Ayala & Sepúlveda, 2018; Mexicanos Primero, 2012).

<sup>11</sup> Isso também envolveu mudanças nos níveis considerados obrigatórios em cada país, que podem ser vistos em profundidade em: Instituto Nacional de Avaliação da Educação (2018). Gaceta de la Política Nacional de Evaluación Educativa. Volume 12. [https://www.inee.edu.mx/wp-content/uploads/2019/01/G12\\_ESP.pdf](https://www.inee.edu.mx/wp-content/uploads/2019/01/G12_ESP.pdf)

Essas últimas interpretações são sustentadas pelo papel desempenhado pelo Banco Mundial, pelo Banco Interamericano de Desenvolvimento, pela UNESCO e pela Agência dos Estados Unidos para o Desenvolvimento Internacional (USAID) no desenvolvimento de sistemas de avaliação na Guatemala, Equador, Panamá, Honduras e Nicarágua, entre outros; bem como o interesse no México para participar nas avaliações internacionais após a adesão à OCDE em 1991 (Martínez Rizo, 2008).

Com relação à participação de países da região em provas internacionais ou regionais, Ferrer e Fiszbein identificam três motivações que se mantêm desde 2003:

- O interesse em comparar-se com os países “pares” da região (no caso da LLECE) e com outros países industrializados (no PISA).
- A oportunidade de fortalecimento técnico do país.
- O impulso de atuar para a melhoria dos sistemas educacionais que decorre do debate sobre os resultados (mesmo quando baixos).

Essas motivações ainda estão presentes nas aplicações mais recentes, como o PISA para o Desenvolvimento de 2018, em que alguns países como o Equador e a Guatemala participaram em grande parte graças ao apoio financeiro que obtiveram do Banco Interamericano de Desenvolvimento e à possibilidade de fortalecer as capacidades técnicas das equipas nacionais (OCDE, 2019). Além disso, os países reconhecem os benefícios econômicos e políticos da adesão à OCDE (Ferrer e Fiszbein, 2015).

Mais adiante, veremos como as diferentes motivações influenciam ou moldam o tipo de uso dado aos resultados das avaliações.

## 2.2 Primeira onda: surgimento de avaliações em larga escala e criação de sistemas de avaliação

A criação de sistemas de avaliação envolveu um grande esforço regional para desenvolver as capacidades de avaliação. Nos anos 90, novas instituições de avaliação foram criadas na região, enquanto as existentes foram fortalecidas. Países como Equador, Cuba, El Salvador, México, Nicarágua, Panamá e Peru desenvolveram áreas de avaliação nos ministérios da educação; na Bolívia e na Colômbia foram criadas novas instituições, e no Brasil

foram ampliadas as funções do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). Outro grupo de países iniciou a avaliação com base na experiência técnica das universidades, formando alianças com o Ministério. Tal foi o caso do Chile, Costa Rica, Guatemala e Honduras (Martínez Rizo, 2008).

A consolidação desses sistemas nos últimos 30 anos tem sido desigual devido a circunstâncias técnicas e políticas:

Alguns países têm mantido seus sistemas funcionando continuamente, mesmo quando ocorrem mudanças de foco ou variações, às vezes acidentais, na institucionalidade dos mesmos. Outros países sofreram descontinuidades significativas e tiveram que começar praticamente do zero em diversas oportunidades ou deverão fazê-lo em um futuro próximo<sup>12</sup> (Ravela, et al., 2007, p. 6).

No nível técnico, as primeiras avaliações desenvolvidas na região se concentraram no aprendizado de idiomas e matemática, usando a teoria clássica das provas e uma abordagem normativa que dá os resultados, na maioria dos casos, em forma de uma pontuação ou número de respostas corretas (Martínez Rizo, 2008).

Durante esse período, houve pouco impacto sobre as políticas educacionais<sup>13</sup>. Diante de uma cultura de avaliação incipiente, observaram-se casos de resultados baixos que deixaram de ser publicados por medo do impacto negativo que poderia gerar (Ravela, et al., 2007). Um exemplo disso foi a aplicação de TIMSS no México, em 1995 (Martínez Rizo, 2008). Experiências semelhantes ocorreram na Bolívia, Equador, Guatemala e Peru (Ferrer, 2006).

### 2.3 Segunda onda: evolução institucional, consideração de fatores associados e feedback de resultados por níveis de realização

Hacia el fin de los noventa y durante los primeros años del siglo XXI, conforme se iban consolidado los sistemas de evaluación, se popularizó el

<sup>12</sup> Para uma descrição mais detalhada e atualizada dos sistemas de avaliação na região, o arquivo sobre as características dos sistemas de avaliação na América Latina estará disponível.

<sup>13</sup> Uma exceção é o Chile, onde o SIMCE foi usado no sistema de seleção de escolas e alocação de fundos.



modelo de evaluación de procesos y comenzaron a utilizarse cuestionarios de contexto para acompañar la interpretación de resultados de aprendizaje (Fortín Morales, 2013; Martínez Rizo, 2008). En este sentido, Cuba y Argentina fueron pioneros en la consideración de factores asociados al aprendizaje. También, otros países ampliaron las asignaturas o campos de conocimiento evaluados.

En el año 2000 comenzó una revisión de las teorías psicométricas utilizadas, y para el año 2008 la mayoría de los países en la región, incluyendo a Bolivia, Argentina, Brasil, Chile, Guatemala y Ecuador, utilizaban pruebas en las que los resultados se reportaban por niveles de logro, en vez de puntajes (Martínez Rizo, 2008)<sup>14</sup>.

No âmbito institucional, os ministérios da educação começaram a assumir funções de avaliação previamente compartilhadas com universidades, e em outros casos, instituições de avaliação foram criadas fora dos ministérios com maior ou menor grau de autonomia (Martínez Rizo, 2008).

Resultados práticos de publicação foram reforçados, embora com uma visão simplista de que esperava ver uma melhoria imediata nos resultados de aprendizagem, eliminando a complexidade dos processos e sistemas educacionais (Martínez Rizo, 2008). O baixo envolvimento do público nas provas gerava pouca discussão sobre os resultados, uma vez publicados (Ferrer, 2006).

## 2.4 Terceira onda: desenvolvimento de medições mais complexas, incorporação de elementos não cognitivos na avaliação e preocupação pelo uso dos resultados

A última década foi marcada por um esforço para responder aos desafios que surgiram da implementação de sistemas de avaliação e da evolução dos sistemas educacionais: expandir o conceito de qualidade educacional, explorar a medição de habilidades interpessoais e desencorajar os usos indesejados, entre outros.

Ferrer e Fiszbein (2015) reconhecem como importantes inovações na região o Índice de Desenvolvimento da Educação Básica (IDEB), do Brasil, e

<sup>14</sup> A Bolívia passou de provas normativas a de critérios em 2000. O Brasil fez o mesmo em 2001 e a Argentina, em 2005. Chile e Guatemala alteraram o feedback dos resultados de pontuação para níveis de conquistas em 2006 e o Equador, em 2008, passou da teoria clássica para a resposta ao item.

o Índice Sintético de Qualidade Educacional da Colômbia (ISCE). Ambos os esforços indicam interesse em usar os resultados das avaliações com informações adicionais<sup>15</sup> que permitam abarcar um conceito mais amplo de qualidade educacional, além de dar seguimento ao longo do tempo e informar as decisões políticas. Uma das críticas dos autores a esses índices é o fato de utilizarem fatores associados simples, em vez de fatores mais complexos, como os encontrados nas avaliações das amostras. Outros países, como Chile e Equador, também avançaram na construção de índices dessa natureza.

De acordo com a recente tendência internacional de reivindicar o valor de habilidades e conhecimentos não acadêmicos para o desenvolvimento pessoal e profissional (Kautz, Heckman, Diris, ter Weel e Borghans, 2014), alguns países da região incorporaram, na última década, elementos de avaliação de aprendizagem não cognitiva. Da mesma forma, a aplicação de questionários de contexto direcionados a estudantes, professores e famílias foi reforçada e aumentada.

A Agência de Qualidade da Educação (ACE), no Chile, incorporou a avaliação Indicadores de Desenvolvimento Pessoal e Social (IDPS): autoestima e motivação escolar, clima e convivência escolar, hábitos de vida saudáveis e formação e participação cidadã (Ayala e Sepúlveda, 2018).

No México, o construto equivalente foi chamado de “habilidades socioafetivas” e refere-se à aprendizagem de saberes que favorecem a convivência escolar (INEE, 2018).

Costa Rica, Equador e República Dominicana também incluem medidas de habilidades socioemocionais em alguns de seus níveis educacionais.

Embora os marcos conceituais para a medição dessas construções, bem como os mecanismos de avaliação, difiram entre os países e ainda estejam em desenvolvimento, espera-se que nos próximos anos eles sejam fortalecidos e expandidos para mais países da região, incorporando avaliações nacionais ou participação em avaliações internacionais em larga escala.

Conforme a limitação formativa das avaliações padronizadas foi ganhando evidência<sup>16</sup>, houve outra alternativa, como parte do mesmo sistema de

<sup>15</sup> El IDEB utiliza resultados de aprendizaje, tasas de aprobación anual y tasas de abandono; el ISCE utiliza rendimiento escolar, aprobación, abandono y clima escolar.

<sup>16</sup> Ver apartado sobre factores que influyen en el uso de resultados y las posturas frente a las

avaliação para fornecer informações em diferentes níveis do sistema educacional. No caso do Chile, Ayala e Sepúlveda descrevem os modelos de avaliações formativas e progressivas da ACE:

Ambos – especialmente o segundo – são para uso interno na escola: não mostram resultados públicos, mas pretendem monitorar o aprendizado ao longo do ano (Ayala e Sepúlveda, 2018, p. 3).

Como será discutido mais adiante, a questão do uso dos resultados é um ponto de interesse e controvérsia nos países da região. A associação de esquemas de incentivos a centros escolares ou professores, bem como as práticas de ranking e uma preocupação com os efeitos homogeneizadores das avaliações padronizadas, levaram a práticas indesejadas e posicionamentos contraditórios em torno das avaliações em larga escala (Jiménez Moreno, 2016; Loza, 2018; Sánchez Cerón e Corte Cruz del Sagrario, 2013).

## 2.5 Marcos institucionais das unidades de avaliação

O marco institucional no qual as avaliações padronizadas são aplicadas é um elemento que influencia o uso de seus resultados de várias frentes: a estabilidade das equipes técnicas encarregadas do projeto e aplicação das provas, a alocação e gestão de recursos humanos e de orçamentos, canais de comunicação entre avaliadores e usuários da informação e a gestão política do discurso em torno das avaliações, entre outros.

As instituições avaliadoras da região podem ser classificadas em dois tipos, de acordo com o marco institucional que as governa: instituições autônomas (ou semiautônomas) e áreas ou unidades dos ministérios da educação<sup>17</sup>.

O relatório do Banco Mundial intitulado *O que aconteceu com os sistemas de avaliação da aprendizagem na América Latina? Lições da última década de experiência* apresenta uma atualização sobre o estado dos sistemas de avaliação em suas estruturas institucionais, na divulgação e uso das informações e na participação em provas internacionais, com foco em cinco países da região: Brasil, Colômbia, Guatemala, México e Peru (Ferrer e Fiszbein, 2015).

---

evaluaciones estandarizadas.

<sup>17</sup> Para una clasificación pormenorizada de los marcos institucionales en cada país, se podrá consultar el fichero de características de los sistemas de evaluación.

Dos países considerados no relatório, México, Colômbia e Peru sofreram mudanças recentes em sua estrutura institucional. Em 2009, foram acrescentadas atribuições ao Instituto Colombiano para a Avaliação da Educação (ICFES), para "oferecer serviços de avaliação, coletar renda e reter excedentes para serem reinvestidos em melhorias técnicas e desenvolvimento de programas" (Ferrer e Fiszbein, 2015, p. 5). Por sua parte, a Direção Geral de Avaliação, Pesquisa e Padrões Educacionais da Guatemala foi legalmente constituída em 2007, embora estivesse em funcionamento há vários anos. No México, o Instituto Nacional para a Avaliação da Educação (INEE) deixou de ser um organismo descentralizado para ser uma entidade autônoma em 2013 (LINEE, 2013). No entanto, essa autonomia foi revogada em uma reforma constitucional de maio de 2019, voltando a ser um órgão descentralizado com uma nova identidade (CPEUM, 2019).

Em 2011, o Chile e o Equador também passaram por importantes reformas nesse sentido, com a descentralização da ACE (Lei 20529, de 2011) e a criação do Instituto Nacional de Avaliação Educacional (INEVAL) como um órgão autônomo (LOEI, 2011), respectivamente.

Embora nos últimos anos houve uma tendência em separar as instâncias de avaliação dos ministérios da educação e conferir a estes maiores poderes, atualmente há um debate sobre esse assunto, que será abordado mais adiante.

**Tabela 4.** Classificação das instituições de avaliação, em nível nacional, por tipo de relacionamento com a autoridade educacional ou outras instâncias.

RELAÇÃO COM AUTORIDADE EDUCACIONAL	PAÍS	INSTITUIÇÃO AVALIADORA
SÃO UNIDADES DO MINISTÉRIO DA EDUCAÇÃO	<b>Argentina</b>	Ministério da Educação, Cultura, Ciência e Tecnologia, Secretaria de Avaliação Educacional
	<b>Costa Rica</b>	Ministério da Educação Pública, Conselho Superior de Educação, Direção de Gestão e Avaliação da Qualidade
	<b>El Salvador</b>	Direção de Avaliação da Aprendizagem do Ministério da Educação

	<b>Guatemala</b>	Direção Geral de Avaliação e Pesquisa Educacional (DIGEDUCA) do Ministério da Educação
	<b>Panamá</b>	Ministério da Educação (MEDUCA)
	<b>Paraguai</b>	Instituto Nacional de Avaliação Educacional (INEE) <sup>18</sup>  Direção Geral de Planejamento Educacional do Ministério da Educação e Cultura
TÉM CERTO GRAU DE AUTONOMIA, MAS COMPARTILHAM RESPONSABILIDADES DE AVALIAÇÃO COM O MINISTÉRIO DA EDUCAÇÃO OU OUTRAS INSTÂNCIAS.	<b>Honduras</b>	Projeto Melhorando o Impacto no Desempenho dos Estudantes em Honduras (Projeto MIDEH)  Direção Geral de Currículo e Avaliação do Ministério da Educação
	<b>México</b>	Instituto Nacional de Avaliação da Educação (INEE) <sup>19</sup>  Secretaria de Educação Pública
	<b>Peru</b>	Ministério da Educação do Peru (MINEDU)  Escritório de Medição da Qualidade da Aprendizagem (UMC)  Instituto Nacional de Estatística e Informática (INEI)
	<b>República Dominicana<sup>20</sup></b>	Instituto Dominicano de Avaliação e Pesquisa da Qualidade Educacional  Direção de Avaliação da Qualidade do Ministério da Educação

<sup>18</sup> Criada na Lei nº 5749, “Que estabelece a Carta Orgânica do Ministério da Educação e Ciência”, de novembro de 2018.

<sup>19</sup> Até 16 de maio de 2019.

<sup>20</sup> A avaliação da primeira infância do ENHOGAR-MICS é realizada pelo Escritório Nacional de Estatísticas.

## 3. Marco conceitual sobre os usos dos resultados das avaliações

Esta seção apresenta uma síntese de referências conceituais sobre o uso dos resultados das avaliações para contextualizar as seções seguintes. Os primeiros modelos focam em tipificar os usos das avaliações. Os últimos abordam a questão das teorias da mudança relacionadas ao uso de avaliações e a melhoria educacional.

### 3.1. Tipologia dos usos da evidência

Carol Weiss propõe uma tipologia de usos da evidência científica, em geral, com o objetivo de esclarecer esse conceito de acordo com o papel que a informação desempenha em uma situação específica. Essa classificação evoluiu entre 1988 (Weiss, 1979) e 1998 (Weiss, 1998) para concretizar-se da seguinte forma:

- **Uso instrumental:** os resultados influenciam a tomada de decisões sobre um programa ou política. Seu uso implica três condições: as descobertas são pouco controversas sem conflito de interesses, as mudanças sugeridas são relativamente fáceis de serem implementadas, o ambiente de implementação permanece relativamente estável. Em casos de paralisia ou emergência, os resultados também podem ser usados para informar as decisões.
- **Uso conceitual:** mesmo sem realizar alterações, os envolvidos modificam e enriquecem sua concepção da situação ou programa. São identificados pontos fortes e fracos que poderão ser considerados quando estejam dadas as condições para atuar.
- **Uso persuasivo:** os resultados são usados para sustentar as decisões tomadas previamente, que requerem legitimidade.
- **Uso "ilustrado":** refere-se aos casos em que os resultados influenciam eventos e instituições para além do escopo da avaliação. Por exemplo, revisões de literatura ou sínteses de conclusões sobre estudos de um mesmo tema ou programa que impulsionam conversas entre acadêmicos, políticos, organizações de lobby, etc. e provocam mudanças na

agenda e nos paradigmas da política, bem como nas opiniões e crenças das comunidades.

Alusões a essa classificação são frequentemente encontradas na literatura da região, especialmente quando há referências de uso instrumental e persuasivo ou simbólico.

Silva, D'Angelo, Fernández, Martínez e Rodríguez (Silva et al., 2012) classificam de maneira mais concreta a forma com a qual os atores educacionais do sistema, no nível macro, acessam e utilizam os resultados das avaliações, com base em um estudo realizado com autoridades educacionais no México:

- Não uso: por desconhecimento, falta de oportunidade no feedback dos resultados ou pouca relevância devido aos níveis muito baixos de desagregação ou aspectos conjunturais de operação ou políticos.
- Uso midiático: são solicitadas informações pontuais sem pesquisar diretamente as fontes, senão terceiros dentro ou fora da organização<sup>21</sup>.
- Uso consultivo: de profundidade variável, as fontes são acessadas diretamente para buscar, corroborar ou contrastar um dado específico ou para um exercício mais amplo de diagnóstico ou planejamento. Há um maior conhecimento das fontes de informação<sup>22</sup>.
- Uso integrado em ações, programas ou políticas: o uso dos resultados visa apoiar ações ou alimentar o projeto de políticas específicas.

Embora a nomenclatura de Weiss permita identificar a intenção do uso da informação, essa classificação é útil para diferenciar com mais detalhes o comportamento de diferentes atores em relação ao gerenciamento dos resultados, a partir de um foco mais operacional.

### 3.2 Objetivos da avaliação

No livro *Para comprender las evaluaciones educativas* (2005), Ravela descreve duas principais classificações de avaliações educacionais, de acordo

<sup>21</sup> Foi o uso predominante entre as autoridades educacionais do mais alto comando entrevistadas no estudo.

<sup>22</sup> Uso observado, principalmente, entre os níveis de gestão intermédia e os órgãos técnicos, bem como as equipes de assessores que participaram deste estudo.

com o tipo de consequências ou finalidade que tenham:

Aquelas que têm consequências diretas importantes para indivíduos ou instituições – os exames para aprovação em um curso, as provas seletivas, as avaliações que definem uma qualificação para os professores ou determinam prêmios em dinheiro – e aquelas cujo objetivo principal consiste em aprender a melhorar, mas que não têm consequências "fortes" (p. 28).

A partir desse agrupamento, o autor desenvolve a seguinte classificação mais detalhada:

- Acreditação ou certificação.
- Ordenamento para seleção.
- Tomada de decisão para a melhoria ("brandas").
- Tomada de decisões "duras" (com consequências de continuidade, ampliação, etc.).
- Estabelecimento de incentivos e sanções.
- Prestação de contas e responsabilização dos resultados.

De acordo com essa classificação, as avaliações de alunos podem ser usadas para a validação ou certificação, ordenamento para seleção, tomada de decisões "brandas" e estabelecimento de incentivos.

As avaliações de professores e de centros educacionais compartilham a mesma variedade de propósitos<sup>23</sup>. Mesmo quando considerados os resultados dos alunos, é possível que uma avaliação em larga escala tenha consequências "brandas" para eles, mas que seja de ordenamento ou incentivos para professores e escolas.

### 3.3 Interação entre visões e funções da avaliação e o uso dos resultados

Fortín Morales (2013) analisa a relação existente entre a visão e a função esperada da avaliação e o desenho e a divulgação dos resultados, com base nas narrativas dos atores educacionais da Guatemala em torno das avaliações padronizadas. Embora o autor advirta que não se trata de um modelo teórico, a classificação que apresenta é útil para analisar a realidade de

<sup>23</sup> É adicionado aos centros o propósito da prestação de contas e responsabilização.



outros países da região.

O autor identifica três visões sobre a avaliação nos diferentes atores envolvidos, tanto no desenho quanto no uso das avaliações:

- Como ferramenta de gestão: a avaliação é mais um elemento para a gestão do sistema como um todo. É um componente de garantia da qualidade. A relação que estabelece com os atores escolares está focada em melhorar o serviço que prestam.
- Como objeto de análise psicométrica: os esforços se concentram na produção de informações tecnicamente sólidas e válidas que possam ser úteis para os tomadores de decisões. Opera sobre o pressuposto de que os tomadores de decisões possuem perguntas que precisam ser respondidas e que coincidem com as levantadas nas avaliações.
- Como fonte de informação pedagógica: interesse em relacionar a avaliação com o que professores e alunos fazem, em vez de considerar o sistema educacional como um todo.

A partir dessas visões, surgem quatro funções esperadas dos resultados na narrativa, que, por vezes, se misturam:

- Controle de qualidade: com um foco na soma e em categorias bem definidas, responde pelo cumprimento ou não dos resultados esperados.
- Exercício de formação: a avaliação é considerada um aprendizado em si e deve provocar mudanças.
- Pesquisa: a avaliação é uma fonte de informação sobre perguntas e outras variáveis relacionadas à qualidade. Delega o uso do conhecimento gerado para outras figuras.
- Orientação pedagógica na sala de aula: enfatiza sua utilidade para o trabalho do professor na sala de aula.
- Ejercicio formativo: la evaluación se considera un aprendizaje en sí mismo y se espera que detone cambios.

**Tabla 5.** Relación entre visiones de la evaluación, funciones de los resultados y la comunicación de resultados

		VISIÓN DE LA EVALUACIÓN			Comunicación de resultados
		Herramienta de gestión	Objeto de análisis psicométrico	Fuente de información pedagógica	
FUNCIÓN ATRIBUIDA A LOS RESULTADOS	Para control de calidad	*	+		Ranking Certificación Acreditación
	Como ejercicio formativo	+		*	Dependerá del actor evaluado
	Como investigación	+	*		Reportes técnicos Reportes de valor agregado Reportes de factores asociados
	Para brindar orientación pedagógica	+	+	*	Sugerencias pedagógicas

\*Uso prioritário de acordo com a visão.

+ Uso admitido, mas não intencional ou dirigido pela visão.

Adaptado de "Evaluación educativa estandarizada en Guatemala: Un camino recorrido, un camino por recorrer (Fortín Morales, 2013).

Mesmo sem se referir intencionalmente a essa classificação, muitos dos trabalhos analisados nas seções a seguir refletem essas visões de avaliação e os usos propostos correspondem às funções descritas aqui.

### 3.4 Teorias da mudança das avaliações em larga escala

No artigo "El uso de los resultados de las evaluaciones en la mejora de los procesos de enseñanza" (Ravela, 2015), o autor identifica seis lógicas de mudança presentes nos diferentes sistemas de avaliação da região, que informam seu desenho e operação:

- Lógica do mercado educacional: as avaliações censitárias e com resultados públicos geram concorrência entre as escolas que incentiva processos de melhoria.
- Lógica de incentivos econômicos diretos: a concessão de incentivos econômicos a partir dos resultados das avaliações censitárias estimula os professores a melhorar suas práticas, contribuindo para uma melhor aprendizagem dos alunos.
- Lógica de pressão social nas escolas: a publicação dos resultados das avaliações gera uma demanda de resultados por parte das famílias e da

sociedade e uma responsabilização dos centros escolares.

- Lógica de desenvolvimento profissional: as avaliações voluntárias e padronizadas com resultados não públicos convidam os professores a revisar suas práticas e o processo de aprendizagem dos alunos, sem a pressão de outras consequências que poderiam desviar o objetivo da melhoria contínua.
- Lógica tecnocêntrica: as avaliações, juntamente com informações sobre o contexto e a escola, oferecem informações valiosas para a tomada de decisões na política educacional ao identificar os fatores com maior incidência nos resultados.
- Lógica da informação estatística: os resultados das avaliações colocam o tema da aprendizagem na agenda pública e da cidadania, sem pretender provocar mudanças diretamente.

Por sua vez, Fortín Morales (2013) descreve quatro lógicas diferentes do impacto das avaliações em três níveis do sistema educacional, de acordo com a função que lhes é atribuída. Essas lógicas estão de alguma forma incluídas e coincidem com as descritas por Ravela.

Níveis de impacto:

1. Sistema: as estruturas administrativas e as ações implementadas por suas instituições.
2. Política: diretrizes e alinhamentos operacionais.
3. Sala de aula: ações na sala de aula como parte do processo de ensino e aprendizagem.

**Tabela 6.** Contribuição da avaliação concebida como uma ferramenta para implementar o controle de qualidade

SISTEMA	AVALIAÇÃO → JULGAMENTO → AÇÃO
POLÍTICA	
SALA DE AULA	

Tabela 7. Contribuição da avaliação concebida como exercício de formação

SISTEMA	AVALIAÇÃO → AÇÃO
POLÍTICA	
SALA DE AULA	AVALIAÇÃO → AÇÃO

Tabela 8. Contribuição da avaliação concebida como pesquisa

	INFORMAÇÃO ADICIONAL ↓
SISTEMA	AVALIAÇÃO → FATORES ASSOCIADOS
POLÍTICA	↘ AÇÃO
SALA DE AULA	↘ AÇÃO

Tabela 9. Contribuição da avaliação concebida como fonte de orientação pedagógica

SISTEMA	AVALIAÇÃO → INTERPRETAÇÃO → COMUNICAÇÃO
POLÍTICA	
AULA	↓ AÇÃO

Tabelas 6-9 Adaptadas de “Evaluación estandarizada en Guatemala: Un camino recorrido, un camino por recorrer” (Fortín Morales, 2013).

Nas seções seguintes, daremos exemplos da aplicação dessas lógicas de mudança e alguns dos resultados que provocaram.

## 4. O uso dos resultados da avaliação em políticas e intervenções educacionais na região

Embora os sistemas de avaliação na região tenham surgido do interesse em medir a qualidade da educação, as lógicas de impacto a serem melhoradas a partir dessas medições e, portanto, seus usos, são muito variadas. Por exemplo, o Chile adotou um foco no controle de qualidade seguindo a lógica do mercado, ponderando os resultados do Sistema de Medição da Qualidade da Educação (SIMCE) como um indicador para a escolha da escola pela família (Ravela, 2015). Na Colômbia, há incentivos direcionados a escolas e professores com base nos avanços apontados no Indicador Sintético de Qualidade, que inclui as pontuações da prova SABER para promover a melhoria de sua prática e a aprendizagem dos alunos (Cabal, Milena e Castro, 2018). Em outros países, como Honduras, a avaliação tanto de professores quanto de alunos não gera consequências e visa fomentar a responsabilização e a tomada de decisões (FEREMA, 2017).

Os usos identificados na literatura da região incluem exemplos de como os resultados da avaliação foram utilizados para:

- Definir a agenda política.
- Relatar exercícios de planejamento estratégico em diferentes níveis do sistema.
- Focar nos programas e intervenções.
- Orientar a definição do currículo e materiais.
- Avaliar ou incentivar os professores.
- Classificar ou gerar um ranking de escolas.

### 4.1 Definição da agenda política

Referindo-se ao discurso político e à agenda educacional, Pedro Ravela ressalta que "é notável que a construção e a difusão de informações sobre a aprendizagem tenham contribuído significativamente para colocar esta última como eixo de preocupações e debates, algo que não acontecia antes dos anos 90" (Ravela, 2015). Rivas (2015), em sua análise dos países da

região após quinze anos de participação no PISA, reconhece que os sistemas de avaliação em larga escala jogaram luz sobre a questão da equidade na educação ao visualizar as lacunas de conquistas existentes entre diferentes segmentos da população.

Pouco a pouco, uma cultura de avaliação ou de políticas baseadas em evidências foi instalada na região, pelo menos no nível discursivo, em que se reconhece a necessidade de se ter avaliações para sustentar o processo de tomada de decisões na política educacional (Ferrer e Fiszbein, 2015; Jiménez Moreno, 2016).

A formalização dos usos dos resultados é muito variada, desde a sua utilização retórica em discursos e como justificativa de decisões tomadas previamente, até intervenções em áreas específicas do sistema.

O exemplo do PISA ilustra esse ponto claramente. A cobertura midiática e retórica do estudo é geralmente usada para impulsionar decisões ou ações do governo já acordadas (Addey e Sellar, 2019; Jiménez Moreno, 2016; Mosqueira e Robinson, 2018). Por outro lado, com base nos resultados do PISA, foram instalados exercícios na região para revisar e repensar o que os alunos estavam aprendendo. A partir daí, surgiram projetos e programas voltados para a revisão dos currículos nacionais, para repensar a formação dos professores e para atribuir novas funções e responsabilidades ao diretor (Rivas, 2015).

No México, os resultados do ENLACE e do PISA foram usados em alguns estados para desenvolver perfis de egresso de estudantes e alinhar a política de avaliação estadual aos padrões federais. Porém, em outras ocasiões, verificou-se que os resultados são utilizados em "diagnósticos" que são realmente justificativas (Silva, D'Angelo, Fernández, Martínez e Rodríguez, 2012). Por sua vez, Loza (2018) propõe que a interpretação contextualizada dos resultados seja acompanhada pela implementação de políticas diferenciadas.

## 4.2 Planejamento estratégico

Quanto ao uso dos resultados nos processos de planejamento estratégico, o ICFES da Colômbia realiza oficinas com as Secretarias de Educação dos departamentos para orientar o uso e a interpretação dos resultados

e usá-los como matéria-prima para o desenho de estratégias dos planos de melhoria, incluindo as etapas de diagnóstico, desenho de ações e seguimento (Mineducation, 2015).

No México, Silva et al. (2012) documentaram casos de uso de resultados de avaliações para a alocação de orçamentos a um nível ou setor específico, e como referência ou dados adicionais em diagnósticos setoriais ou programáticos mais amplos. No Chile, a Comissão para o Desenvolvimento e Uso do SIMCE relatou o amplo uso dos resultados do SIMCE no projeto e na avaliação de políticas como indicadores de metas nacionais, bem como instrumentos para avaliar programas e intervenções, tanto nacionais quanto de escolas (Sandoval-Hernández, 2016).

### 4.3 Direcionamento de programas e intervenções

Há menos informações sobre o uso concreto dos resultados das avaliações em programas e intervenções específicas, provavelmente porque essa informação seja encontrada em documentos internos ou regulamentos de programas. Algumas experiências específicas do México mostram que os resultados do ENLACE foram utilizados como critério de participação e instrumento de prestação de contas do programa Escolas da Qualidade (Silva et al., 2012). Da mesma forma, as diretrizes para o Sistema de Assessoria e Acompanhamento às Escolas (SISAAE) referem-se aos resultados das avaliações externas e internas para identificar as necessidades pedagógicas e definir objetivos e metas para o ciclo escolar (CNSP, 2019).

No Chile, o próprio Ministério da Educação indica que os resultados das avaliações podem ser utilizados para focalizar estabelecimentos que podem se beneficiar de intervenções, apoios ou programas como o “900 Escuelas”, ou receber fundos para projetos de melhoria (Sandoval-Hernández, 2016; Rivas, 2015). No Uruguai, os resultados foram utilizados para manter e expandir o programa de escolas de tempo integral, e em Bogotá, para tomar decisões sobre investimentos em infraestrutura e refeições escolares (Rivas, 2015).

### 4.4 Orientação do currículo e materiais

No México, foi registrado um caso em âmbito federal do uso dos resultados da aprendizagem para orientar a decisão sobre os temas a serem

reforçados nos materiais educacionais, ajustes no currículo e desenho de cursos (Silva et al., 2012). Bogotá, Paraná e São Paulo<sup>24</sup> incluem em seus sistemas de avaliação docente um componente de formação continuada que aborda como melhorar o desempenho dos alunos em provas padronizadas (Rivas, 2015). No Chile, as provas padronizadas fomentaram um apego ao currículo estabelecido (Sandoval-Hernández, 2016) e, em geral, observou-se uma influência do PISA na definição de padrões de aprendizagem na região, embora essa influência também enfrenta outras dinâmicas internas dos países, tais como pressões sindicais para manter postos de trabalho ou demandas sociais para priorizar certos saberes (Rivas, 2015).

#### 4.5 Incentivos a docentes

Segundo o estudo do PREAL, os resultados de provas padronizadas são interpretados como “um indicador da qualidade do trabalho dos professores e um critério para a concessão de incentivos econômicos” (Ravela et al., 2007, p. 14). Alguns países, como Colômbia e México<sup>25</sup>, estabeleceram incentivos ligados aos resultados das avaliações por meio de programas como o Incentivos à Qualidade Docentes, no primeiro caso, e o Estímulos da Qualidade Docente, no segundo (Silva et al., 2012; Cabal, Milena e Castro, 2018). De maneira informal, no Chile, houve relatos de pagamentos a professores cujos alunos obtiveram melhores pontuações, apesar das advertências da própria Agência de Qualidade da Educação contra esse tipo de prática (Sandoval-Hernández, 2016).

#### 4.6 Classificação ou classificação das escolas

Por fim, deve-se mencionar o uso da classificação ou ranking de escolas para os seguintes fins:

- i) fomentar a responsabilidade das escolas pelos seus resultados; ii) fornecer informações às famílias para apoiar a escolha da escola de seus filhos; iii) promover a competição entre os centros educacionais para obter melhores resultados.

Utilizar os resultados de avaliações padronizadas para fornecer

<sup>24</sup> Paraná e São Paulo operam sistemas de avaliação subnacionais.

<sup>25</sup> No México, a prática foi suspensa em 2013 com a substituição do ENLACE com o Plano Nacional de Avaliação de Aprendizagem (PLANEA).



incentivos econômicos às escolas que obtenham melhores resultados ou melhorias em relação aos seus resultados em avaliações anteriores (Ravela, et al., 2007, p. 14).

É o caso do Chile, como mencionado no início, embora nos últimos anos tenham sido feitas tentativas de afastamento da prática do ranking para enfatizar o uso formativo das avaliações (Ayala e Sepúlveda, 2018; EDUY21, sf). Na Guatemala, os resultados são dados pelas escolas e ordenados de acordo com os resultados dos alunos, considerando o nível socioeconômico (Martínez Rizo, 2008). Embora no México não haja consequências “duras” para as escolas, a publicação dos resultados e rankings elaborados pelos meios de comunicação ou pelas próprias autoridades estatais pressiona os supervisores, diretores e equipes docentes ou gera práticas de simulação (Sánchez Cerón e Corte Cruz del Sagrario, 2013). No caso de Honduras, os rankings não são emitidos como tais, senão que se publicam os resultados desagregados no nível municipal para promover a responsabilização e a concorrência saudável entre departamentos e municípios, com resultados encontrados de legitimação e danos aos centros educacionais (FEREMA, 2017). O Uruguai foi o primeiro país a publicar resultados estabelecendo comparações entre escolas de nível socioeconômico similar, e no Brasil cada escola é comparada a um grupo de escolas anônimas que compartilham características similares, de acordo com o IDEB (EDUY21, sf; Rivas, 2015).

## 5. Fatores que influenciam o uso (ou o não uso) dos resultados das avaliações

Esta seção inclui trabalhos que documentaram fatores que influenciam o uso dos resultados das avaliações mencionados anteriormente. A busca se centrou, na medida do possível, em referências específicas ao uso dos resultados de avaliações em larga escala, diferenciando-os de indicadores ou outras informações resultantes dos processos de avaliação. Existem vários estudos que recuperam informações gerais sobre esses fatores de uso dos resultados, mas foram encontrados poucos estudos que descrevem detalhadamente, em nível local ou nacional, as nuances e dinâmicas do uso dos

resultados por diferentes atores<sup>26</sup>.

Embora o uso dos resultados, na realidade, dependa da combinação de múltiplos fatores que nem sempre podem ser separados, para fins de análise, as referências foram agrupadas em quatro categorias:

- Fatores políticos: referem-se às realidades do plano político-governamental, bem como à interação entre as políticas do sistema educacional e seus atores.
- Elementos institucionais: aqueles que se referem à estrutura institucional na qual as avaliações são realizadas e utilizadas.
- Aspectos técnicos: relacionados, seja com elementos técnicos das próprias provas que influenciam os usos, como as capacidades técnicas para utilizá-los.
- Comunicação e divulgação: referem-se a diferentes práticas de registro e divulgação dos resultados.

## 5.1 Fatores políticos

Segundo a análise de Rivas (2015), durante os primeiros quinze anos do novo milênio, houve uma tendência regional em renovar a governabilidade centralizada dos sistemas educacionais. Nesse contexto, as avaliações com foco na medição da qualidade tornaram-se uma ferramenta de gestão de resultados e de legitimação da transparência dos Estados e governos.

Um fator relacionado a essa legitimidade, que ganhou peso no cenário político da região, é o tipo de cobertura da mídia atribuída aos resultados das avaliações. A publicação dos resultados das avaliações foi um grande passo em termos de transparência e prestação de contas dos governos, conforme reconhecido pelo estudo PREAL:

Parece ter sido superado o estágio em que as autoridades de vários países interferiam na divulgação dos resultados das avaliações quando não lhes pareciam favoráveis. Em geral, há maior transparência na região na divulgação dos resultados (Ravela, et al., 2007, p. 6).

<sup>26</sup> Foi identificado um estudo realizado em 2012 no México e outro em 2013 na Guatemala. Axel Rivas recolhe exemplos em seu livro América Latina depois do PISA: Lições aprendidas da educação em sete países (2000-2015).

No entanto, o início da publicação dos resultados da avaliação teve complicações. Caracterizou-se pela cobertura sensacionalista, alarmista e desqualificante de governos e sistemas educacionais (Ferrer e Fiszbein, 2015), apesar dos esforços dos órgãos de avaliação para descartar essas práticas.

Embora a própria OCDE recomende contextualizar os resultados dessas provas nas condições socioeconômicas e culturais de cada país para evitar interpretações reducionistas (OCDE, s/f), é evidente que as listas nas quais os países participantes são classificados (ranking) constituem atualmente o motor do efeito midiático adquirido em quase todos os países, seja para questionar o mal desempenho de seus respectivos sistemas educacionais ou para exaltar suas realizações (Jiménez Moreno, 2016, p. 6).

Vários autores concordam que a interpretação dos resultados das avaliações é simplista, não se aprofunda no aspecto medido pelas provas nem considera as dinâmicas e complexidades do sistema (Jiménez, 2017; Ravela, et al., 2007).

Em suma, o relatório do Banco Mundial (Ferrer e Fiszbein, 2015) reconhece o progresso na educação da mídia com relação ao uso dos resultados:

Os meios de comunicação e os jornalistas estão cada vez mais informados sobre as questões educacionais; portanto, agora eles podem ir além das manchetes sensacionalistas que eram tão frequentes e populares no passado. As más notícias ainda são material para manchetes negativas e provocam duras críticas às políticas governamentais, mas, ao mesmo tempo, parecem contribuir para novos debates sobre a qualidade da educação em diversos fóruns (Ferrer e Fiszbein, 2015, p. 12).

O relatório do Banco Mundial postula que "provavelmente, como resultado dessa cobertura massiva, os políticos estão prestando mais atenção à importância e ao protagonismo dos resultados das avaliações de aprendizagem, embora não se sintam necessariamente responsáveis por esses resultados" (Ferrer e Fiszbein, 2015, p. 12). No entanto, lamentam que os políticos sucumbam à pressão e "tendam a tomar decisões precipitadas para demonstrar que estão tomando medidas para mudar a situação" (Ferrer e Fiszbein, 2015, p. 13). Às vezes, essas decisões levam a programas que não consideram adequadamente a dinâmica e os processos do próprio

sistema educacional e seu impacto nos processos de ensino e aprendizagem. Portanto, tornam-se atalhos que não dão os resultados esperados (Ravela, et al., 2007; Rivas, 2015).

Da mesma forma, é difícil reorientar os discursos estabelecidos na mídia, resultado de usos históricos dos resultados, para promover um uso das avaliações sem uma abordagem punitiva ou competitiva (Ayala e Sepúlveda, 2018; Ravela, 2015).

Harvey Sánchez (2018) ressalta que, às vezes, os resultados são descartados ou parcialmente utilizados, por interesses alheios à educação. Ferrer (2006) reconhece o mesmo ao incluir a vontade política como um dos fatores que influenciam o uso de resultados na região. Como exemplo, podemos citar o caso da Bolívia, que em 2000 interrompeu a aplicação da prova SIMECAL e deixou de participar das provas do LLECE por razões ideológicas e políticas. Até a presente data, o país não retomou um sistema de avaliação em larga escala, embora tenha se reincorporado à LLECE para a aplicação do ERCE 2019 (Urquiola, 2011; Addey e Sellar, 2019).

Tanto Addey (2019) quanto Mosqueira e Robinson (2018) abordam o caso oposto, no qual os resultados das avaliações servem como alavanca ou justificativa para gerar um senso de urgência, chegando ao ponto de criar realidades que facilitam os governos a tomar decisões difíceis ou controversas.

Do ponto de vista da política educacional, referindo-se ao caso das avaliações no Uruguai, Ravela (2015) apresenta a probabilidade de que as avaliações voluntárias online com alto potencial pedagógico sejam utilizadas principalmente por professores comprometidos e motivados. Seguindo essa mesma linha, Rivas (2015) se refere à sua experiência e convida a estudar o fenômeno pelo qual, em ambientes com baixas consequências ou consequências “brandas”, os professores não se esforçam para entender os resultados das avaliações. Em ambos os casos, argumenta-se que a iniciativa do professor é insuficiente para instalar uma cultura de uso de resultados no sistema.

## 5.2 Elementos institucionais

Conforme estabelecido na primeira seção do documento, existem vários modelos institucionais na região para gerir os sistemas de avaliação.

Sánchez (2018) defende que as instituições avaliadoras devem ter autonomia e independência que as distanciem da pressão dos poderes externos. Embora sejam reconhecidas as vantagens da objetividade e da estabilidade técnica em se ter instituições autônomas independentes dos ministérios da educação, observou-se que essa separação dificulta o fluxo de informação, a vinculação e a apropriação de resultados pelos usuários (Fortín Morales, 2013; Ferrer e Fiszbein, 2015). Dito isto, Ferrer e Fiszbein não encontraram diferenças substanciais nesse sentido nos países onde as unidades de avaliação fazem parte do ministério.

Isso sugere que, para além das relações institucionais, o peso das relações informais e pessoais desempenha um papel importante no uso dos resultados pelas autoridades educacionais. A esse respeito, Silva et al. (2012) encontraram o seguinte:

No caso das autoridades federais, duas condições impactaram os usos predominantes. Por um lado, as particularidades que delimitam as competências de cada direção geral e, por outro, a proximidade e o contato direto entre essas autoridades e o próprio INEE (2012, p. 55).

Os sistemas de avaliação devem ser concebidos como redes e definir claramente como interagem no sistema educacional em geral (Fortín Morales, 2013). Essas relações nem sempre são claras, independentemente de a agência avaliadora estar dentro ou fora do ministério (Fortín Morales, 2013; Ferrer e Fiszbein, 2015).

Pensamos na avaliação de maneira excessivamente independente, não como parte de um conjunto mais amplo de ações e elementos da política educacional (...). Pensamos na avaliação da aprendizagem em larga escala independentemente das avaliações das aprendizagens que ocorrem diariamente nas salas de aula e, em muitos casos, das mudanças curriculares. Precisamos pensar na avaliação como parte e articulada a um conjunto de ações específicas da política educacional (Ravela, 2015, p. 80).

Essa falta de articulação frequentemente pesa sobre o grau de apropriação da informação pelos usuários. Nesses casos, a barreira entre os fatores políticos e institucionais não é clara, especialmente quando há conflitos de prevalência de uma área ou instituição sobre outra, ou quando se percebe rivalidade nas informações relatadas. É necessária uma documentação

adicional dessas situações, mas Silva et al. (2012) encontraram informantes na educação secundária superior que manifestaram alguma suspeita quando não foram consultados antes da publicação de números que discordavam de seus próprios dados. Na mesma linha, o estudo de Ferrer (2006) indica que a participação e o envolvimento dos usuários desde as etapas do desenho das avaliações podem favorecer essa apropriação e consequente uso dos resultados. Fortín Morales (2013) insta a considerar de maneira explícita e clara as estruturas operacionais do sistema educacional ao pensar em avaliação.

Outro efeito da falta de clareza na articulação entre sistemas de avaliação e outras áreas do sistema educacional é a dificuldade em atribuir responsabilidades para as ações de melhoria. As recomendações emitidas a partir da avaliação não são vinculativas ou não há clareza sobre os fatores subjacentes aos resultados e quem é o responsável por executar as correspondentes ações de melhoria (Silva et al., 2012; Fortín Morales, 2013). Vários autores recomendam e identificam áreas do sistema educacional que deveriam ser articuladas em torno do desenho de políticas, com base nas carências reveladas pelas provas de aprendizagem; por exemplo: avaliação, desenvolvimento curricular, formação inicial e desenvolvimento profissional de docentes (Ravela et al., 2007).

### 5.3 Aspectos técnicos

Embora exista comunicação adequada entre as equipes de avaliadores e os operadores do sistema educacional, ainda há falta de clareza na região sobre os propósitos e fins das avaliações (Mosqueira e Robinson, 2018). Isso é atribuído, por um lado, à falta de consenso sobre os próprios objetivos e padrões da educação e do currículo, o que afeta a definição do que deve ser avaliado e com que finalidades. Por outro lado, poucos atores entendem que o uso dos resultados é claro e definido a partir do momento do desenho técnico da avaliação, e não o contrário; e que uma avaliação não pode servir para todos os propósitos (Ravela, et al, 2007; Ravela., 2005):

Sempre que ambas as finalidades [consequências "fortes" e foco de formação] estão presentes no mesmo ato de avaliação, as consequências fortes prevalecem sobre os objetivos da formação e a distorcem (Ravela, 2015, p. 103).

Este desconhecimento técnico sobre as avaliações, em algumas ocasiões, faz com que a urgência dos tempos políticos atrepele os tempos necessários para estabelecer propósitos claros e garantir a qualidade técnica das avaliações Ravela, et al., 2007).

Com frequência, há contrapontos entre visões e argumentos com relação ao objetivo de avaliações que geram insatisfação nos usuários: os resultados gerados com fins de qualidade são criticados por sua falta de foco na formação; ou o rigor da pesquisa não necessariamente atende as necessidades de orientação pedagógica (Fortín Morales, 2013).

O comentário de uma autoridade educacional no México ilustra este ponto:

Eu gostaria de obter mais ajuda do INEE para fazer esse tipo de análise de conjuntura em matéria de política. Não a temos porque atuam sob outra dinâmica, mais acadêmica, precisam de tempo para analisar o problema, levantar e comprovar as hipóteses. Temos dinâmicas muito diferentes em relação à análise da informação e seu uso (...) (Silva et al., 2012, p. 69).

Em relação à tensão entre a lógica e os tempos das unidades de avaliação e os das áreas operacionais do sistema educacional, alguns países como Peru, Chile e México preferiram apostar em levantamentos periódicos e regulares, com uma distância no tempo suficiente para permitir processar e entregar resultados em tempo hábil, sem sacrificar sua qualidade técnica (Ferrer, 2006; Silva et al., 2012).

Dada a expectativa de usar os resultados de avaliações padronizadas para informar políticas e intervenções, surgem as próprias limitações dos levantamentos. Os resultados de avaliações em larga escala, especialmente quando são censitários, dificultam a coleta de fatores do contexto escolar, social e do próprio sistema, necessários para contextualizar adequadamente os resultados em cada país ou localidade e projetar intervenções apropriadas (Jiménez Moreno, 2016; Ferrer e Fiszbein, 2015). Os resultados das avaliações, por si só, podem identificar um problema, mas não dão pistas sobre sua possível solução. Sandoval-Hernández (2016), também citado por Jiménez Moreno (2016), argumenta que os resultados do PISA podem ser usados para sustentar todos os tipos de políticas, até mesmo

contraditórias entre si.

Além disso, essa limitação gera interpretações simplistas e descontextualizadas que são amplamente criticadas na literatura. O relatório PREAL relata que, em 2007, ainda persistiam as práticas de divulgação de resultados regionais, do setor escolar ou de escolas, sem informações sobre as condições e características do contexto, o que levava a interpretações isoladas ou errôneas (Ravela et al., 2007). Jiménez Moreno (2016) retoma de Backhoff um exemplo do México, onde a conversa midiática não considerou em sua interpretação sobre as tendências nos resultados do PISA a inclusão de estudantes no sistema educacional que, antes, não frequentavam a escola. Dessa forma, os resultados mostram uma aparente estagnação do sistema educacional.

Referindo-se às práticas existentes em 2006, Martínez Rizo comentou:

Hoje, as aplicações proliferam e começam a cansar as escolas; os resultados são debatidos, às vezes com pouco conhecimento de sua complexidade. Para que as perspectivas favoráveis se concretizem, será necessário que uma visão objetiva das possibilidades de avaliação prevaleça sobre as visões superficiais (2008, p. 26).

Rivas compartilha seu posicionamento lamentando a proliferação de educadores inexperientes que opinam sobre a educação com base nos resultados das avaliações, e que geraram uma “onda de desgaste sobre os professores” (Rivas, 2015, p. 98).

O relatório do Banco Mundial de 2015 ainda menciona como um desafio pendente o aprofundamento no debate sobre os resultados das avaliações (Ferrer e Fiszbein, 2015).

Em resposta às inquietações sobre a contextualização adequada dos resultados, Ravela et al. (2007) propõem a realização de estudos de valor agregado, mas reconhecem a falta de capacidade técnica na região nessa área.

Referindo-se especificamente ao uso dos resultados, Fortín Morales (2013) reconhece que há pouca capacidade técnica para utilizar os resultados em sala de aula, e não está claro quem deve desenvolver essa capacidade. Ravela et al. (2007) coincidem, especificando que a maioria dos professores desconhece o que as provas medem e, portanto, não sabem



como usar os resultados adequadamente. Ferrer e Fiszbein (2015) acrescentam como barreira o fato de os professores e as famílias não estarem preparados para ler os formatos em que os resultados das avaliações são apresentados (barras, gráficos, escalas, tabelas comparativas, textos descontínuos), apesar dos esforços das instâncias de avaliação em simplificar as informações e os relatórios.

O relatório PREAL relata uma baixa capacidade das equipes técnicas dos ministérios para responder aos problemas identificados nas avaliações e, exceto no caso do Brasil e da Argentina, Rivas identificou uma falta de atenção à formação técnica de funcionários do sistema educacional (Ravela et al., 2007; Rivas, 2015). Como agravante do problema, a alta rotatividade das equipes nos ministérios da educação dificulta o fortalecimento do conhecimento e da gestão das avaliações, bem como o vínculo entre instâncias de avaliação e operacionais (Silva et al., 2012).

Embora o objetivo desta revisão não seja abordar detalhadamente os aspectos técnicos das provas, a seguir é apresentado um esquema da relação estabelecida pelos autores do estudo do PREAL, entre o nível de qualidade técnica das provas, a divulgação dos resultados e os usos e práticas que derivam deles. O conteúdo da tabela resume e explica, em grande parte, os fenômenos observados na região em relação ao uso dos resultados.

**Tabela 10.** Relação entre qualidade técnica e divulgação das provas

DIVULGAÇÃO E USO	QUALIDADE TÉCNICA	
	BOA	RUIM
Divulgação e uso Adequados às características técnicas da prova	Contribuição. Possibilidade de impacto na melhoria do desempenho nas áreas avaliadas.	Efeitos perversos, interpretações e decisões erradas.
Não adequados às características técnicas da prova	Efeitos perversos, interpretações e decisões erradas.	Efeitos perversos, interpretações e decisões erradas.
Escassos ou nulos	Desperdício de recursos e oportunidades.	Desperdício de recursos.

Extraído de “Las evaluaciones educativas que América Latina necesita” (Ravela et al., 2007, p. 10).

## 5.4 Comunicação e divulgação

Por fim, deve-se mencionar o papel desempenhado pelos mecanismos de produção de relatórios e divulgação dos resultados sobre seu uso.

O estudo do Banco Mundial por Ferrer e Fiszbein (2015) reconhece um avanço significativo durante os últimos dez anos nas estratégias de comunicação de resultados para diferentes públicos, embora ainda haja muito por fazer. O mesmo relatório identifica as seguintes maneiras de apresentar os resultados e sugestões para sua interpretação:

- Relatórios completos com resultados nacionais e subnacionais.
- Relatórios de progresso longitudinais.
- Relatórios individuais para estudantes e escolas (em contextos de altas consequências).
- Bases de dados.
- Análise secundária dos dados.
- Recomendações pedagógicas a partir dos itens liberados.
- Mecanismos de busca online para encontrar resultados e exemplos de conteúdos avaliados.
- Apresentações audiovisuais em diferentes formatos.

Segundo Fortín Morales, os desafios da divulgação devem considerar que:

a) a informação chegue aos usuários; b) a informação seja representada de uma maneira que os usuários possam compreender; c) a informação fornecida seja a necessária para iniciar mudanças em favor da qualidade educacional (2013, p. 17).

O mesmo autor reivindica uma comunicação explicitamente voltada para os níveis operacionais do sistema educacional.

Nesse sentido, existe uma consciência generalizada sobre a necessidade de diversificar o tipo de produtos para se referir a diferentes públicos e várias iniciativas já estão em andamento para abordar esse tema. Ayala e Sepúlveda (2018) destacam os esforços recentes que foram feitos no Chile para escrever relatórios, resumos de políticas e relatórios que forneçam apoio aos tomadores de decisão. Bracho (2018) reconhece que a oferta de informação do INEE, no México, ainda é muito uniforme e pouco diferenciada, enquanto outros estudos constatam que, antes de 2013, havia poucos canais de distribuição de relatórios e resultados em nível estatal no país (Silva et al. al., 2012).

Em geral, percebe-se que os países com sistemas de resultados de consequências altas ou "fortes" dedicam maiores esforços à sua divulgação (Ferrer, 2006).

Não apenas é necessário diversificar os formatos de divulgação, mas também atender à sua qualidade técnica e comunicativa para o público ao qual está direcionada. Vários países da região procuraram inovar o conteúdo dos relatórios às escolas para torná-los mais didáticos e simples; a inclusão da descrição dos níveis de realização apresenta um potencial de uso pedagógico cujos resultados ainda não foram documentados (Ferrer e Fiszbein, 2015). Em Honduras, verificou-se que os relatórios publicados em âmbito municipal geraram maior apropriação e resposta do que os publicados com resultados em âmbito nacional (FEREMA, 2017).

Além disso, em toda a região, foram feitos esforços para educar o público usuário dos resultados das avaliações (Ferrer e Fiszbein, 2015). No Chile, por exemplo, são realizados seminários e oficinas com professores, diretores e atores da opinião pública para colocar temas de discussão em torno dos resultados das avaliações (Ayala e Sepúlveda, 2018). Na Colômbia, o ICFES entrega relatórios às Entidades Territoriais Certificadas (ETC) com os resultados de todas as escolas filiadas, com um guia para a interpretação e uso dos resultados, e foram realizadas iniciativas de oficinas para discutir os resultados nas escolas (Rivas, 2015). O Uruguai realiza oficinas com profissionais da educação para discutir os resultados e fornece guias de uso para as escolas (Rivas, 2015).

Os sites dos institutos ou unidades de avaliação não podem deixar de ser mencionados. A revisão de Ferrer e Fiszbein (2015) constata que a maioria deles não está bem desenhada e, geralmente, apresenta falhas técnicas. A estrutura das páginas não tem um objetivo claro de comunicação e é difícil encontrar os relatórios, dados ou relatórios. As informações se repetem e os nomes das ligas não são intuitivos para o público não especialista. As páginas do INEVAL, do Equador, e a do INEE recentemente atualizada<sup>27</sup>, são destacadas como exceções por estarem explicitamente endereçadas a diferentes públicos (INEVAL, 2019; INEE, 2019).

---

<sup>27</sup> Durante a elaboração deste trabalho, verificou-se que os portais da internet são frequentemente alterados com as mudanças na administração, por vezes, descartando os esforços anteriores que facilitavam o acesso para os usuários.

## 6. Posicionamentos quanto às avaliações padronizadas

Esta última seção da revisão busca recuperar as diferentes vozes que surgiram na conversa sobre os resultados das avaliações na região. Opiniões, críticas, defesas e algumas recomendações são apresentadas, agrupadas em duas categorias principais:

- Teórico-ideológica: trabalhos que abordam, a partir de posições críticas, as definições e marcos conceituais das avaliações em larga escala no âmbito dos sistemas de avaliação para medir a qualidade educacional; bem como questionamentos sobre as teorias da mudança subjacentes ao uso dos resultados das avaliações.
- Técnico-instrumental: referências a questionamentos e reflexões derivadas da implementação das avaliações em larga escala e seus efeitos sobre os sistemas educacionais da região.

### 6.1 Teórico-ideológica

O uso do termo “qualidade”, como referência para a avaliação, tem sido fortemente criticado devido à falta de consenso em relação à definição de qualidade educacional.

Não existe um posicionamento único sobre o que se entende por “ideal educacional” e, portanto, por “qualidade”, pois isso depende da corrente pedagógica e ideológica que um sistema educacional adota em determinado momento (Mosqueira e Robinson, 2018; Fortín Morales, 2013). Casassus (2007) denuncia que, diante dessa lacuna, o termo “qualidade” na educação foi reduzido a uma abordagem neoliberal voltado para servir o mercado de trabalho e, de forma operacional, ao que pode ser medido em provas padronizadas. Essa visão é compartilhada por Jiménez Moreno (2016), que critica a influência do PISA na definição da qualidade da educação com base no domínio de habilidades e competências para a inserção no mercado de trabalho.

Prats (2011) aponta que instrumentos como o PISA foram constituídos em

avaliações do capital humano de cada país no contexto da competição internacional, e que essa associação pode influenciar negativamente a finalidade mesma da educação.

Nesse sentido, a crítica mais forte recai sobre sua influência, não apenas em termos de definição de referências educacionais para os países, mas também em estabelecer uma agenda global mais ampla para o desenvolvimento e, concretamente, para a educação (Casassus, 2007; Jiménez, 2017; Jiménez Moreno, 2016; Sánchez Cerón e Corte Cruz del Sagrario, 2013).

Além da discussão em torno do termo "qualidade", mas relacionada à importância de alcançar definições teóricas no nível do sistema educacional, os autores do relatório do PREAL (Ravela et al., 2007) argumentam que a falta de clareza nas metas e padrões de aprendizagem, bem como os perfis de egresso dos estudantes em diferentes níveis, dificultam não apenas a avaliação em si, mas também o ensino. Assim, reforçam a advertência dos riscos que envolvem a comparação entre provas padronizadas com medidas da qualidade da educação ou do sistema educacional.

Além deste último ponto, Ravela destaca a importância de ampliar a própria noção de "resultados", pois:

Os "resultados" da avaliação educacional não são iguais aos resultados do trabalho educacional (...) Os "resultados" – ou produtos – da avaliação devem incluir uma visão e uma avaliação do conjunto da situação educacional, bem como a análise dos processos que ocorrem no sistema educacional e que estão na base de suas realizações e insuficiências (Ravela, 2015, p. 84).

Nesse sentido, começam a surgir como alternativas medidas complementares e índices que considerem variáveis adicionais (Bracho, 2018). Como mencionado anteriormente, o Brasil e a Colômbia estão liderando o caminho com o IDEB e o ISCE, respectivamente (Ferrer e Fiszbein, 2015).

Outra crítica a essa abordagem de qualidade nas avaliações em larga escala é que ela ignora o restante dos componentes e recursos, as características, trajetórias e outras habilidades dos estudantes, bem como os processos e limitações do sistema educacional e a própria responsabilidade do Estado para com a tarefa educacional (Mosqueira e Robinson, 2018; Jiménez

Moreno, 2016; Ravela, 2015).

Algumas consequências são a demarcação do Estado quanto ao seu papel educativo, a transferência do peso da responsabilidade da aprendizagem aos professores e do controle das escolas às famílias, além da tentativa de suplantar as funções de acompanhamento e atualização que correspondem aos supervisores ou outras figuras que promovam a abordagem “formativa” das provas (Ravela, 2015; Jiménez, 2017).

Nesse sentido, o estudo do PREAL também alerta para o perigo de que, por meio de sistemas de ranking ou políticas de incentivos docentes com base em resultados, o Estado transfira sua responsabilidade educacional para escolas, professores e famílias. Nesse cenário, o Estado se limita a avaliar e informar, deixando de cumprir sua obrigação de desenvolver capacidades para a melhoria da educação e de priorizar os contextos mais desfavorecidos (Ravela, et al., 2007).

Tanto Casassus (2007) quanto Jiménez (2017) e Ravela (2015) denunciam a culpabilização dos professores pelos baixos resultados; e o último alude aos resultados de um estudo no Uruguai, no qual os professores sentem a exigência de resolver problemas sociais e de desenvolvimento econômico através de sua função educadora, enquanto não recebem as ferramentas apropriadas e sua profissão não é valorizada. Exemplos dessa pressão também foram relatados no Brasil, Chile e México (Sánchez Cerón e Corte Cruz del Sagrario, 2013).

## 6.2 Técnico-instrumental

Desde a sua criação, os sistemas de avaliação geraram grandes expectativas como recursos que permitiriam projetar políticas e intervenções para a melhoria com base em evidências. Quase 30 anos após o início de sua implementação, fica claro que as avaliações de aprendizagem, por si só, não desencadeiam processos de melhoria.

Existem algumas críticas ao modelo de avaliação de processos, argumentando ser útil como modelo explicativo, mas que não informa como gerar as condições para melhoria. Embora, em geral, a pesquisa sobre fatores associados às conquistas seja valorizada, alguns teóricos defendem que não há clareza sobre como interpretar a causalidade (Fortín Morales, 2013).

Casassus (2007) expressa sua decepção diante das expectativas geradas pelas avaliações e dos poucos resultados de melhoria na região, considerando o investimento que se faz para realizar essas provas. Em seu posicionamento, propõe a supressão das provas e que esses recursos sejam destinados diretamente à melhoria do sistema.

Em resposta, outras posições defendem a necessidade de informações sobre as conquistas de aprendizagem, reconhecendo sua função como mais um elemento dentro de um sistema integral de avaliação e dos próprios sistemas educacionais (Mosqueira e Robinson, 2018; Martínez Rizo, 2008). A partir da experiência na Bolívia, onde não se aplicam avaliações padronizadas desde o ano 2000<sup>28</sup>, Urquiola (2011) expressa as dificuldades em avaliar a qualidade da educação e o efeito do investimento no setor educacional sem dados sobre as conquistas educacionais, e defende a reinstalação das provas em larga escala. Sánchez (2018) alerta para os perigos de desenvolver políticas sem evidências, onde prevalecem os argumentos do governo baseados em histórias ou retóricas, e onde as ações dependem dos promotores de políticas.

Buscando um ponto de encontro, o relatório do PREAL expressa o seguinte:

A avaliação é apenas um dos vários elementos da política educacional que devem ser alinhados adequadamente: a formação docente inicial e em serviço, a carreira e condições para o trabalho docente, a gestão escolar, a supervisão, a elaboração dos currículos, os livros e materiais educacionais, um investimento de recursos proporcional às necessidades das diversas populações e uma ação determinada pelos responsáveis pela educação para solucionar os problemas detectados, entre outros (Ravela et al., 2007, p. 5).

O texto reflete claramente a complexidade de se articular os processos de melhoria e oferece uma perspectiva sobre o papel viável e as limitações das próprias avaliações padronizadas.

---

<sup>28</sup> Desde sua criação em 1996, o Sistema de Medição da Qualidade (SIMECAL) aplicou provas em 1997 e 2000, até seu desmantelamento em 2004 (Martínez Rizo, 2008). Em 2017, o país retomou sua participação nos estudos da LLECE com a aplicação do estudo preliminar denominado Pós-TERCE, coordenado pela LLECE e pelo Ministério da Educação.

Em termos de políticas públicas, embora os resultados dessas provas sejam úteis para realizar um diagnóstico, Jiménez Moreno sustenta que é insuficiente informar o desenho das políticas educacionais:

(...) porque, para realizar essa tarefa, é necessário considerar um grande número de aspectos do contexto escolar e social que, devido às suas próprias limitações, as provas não são capazes de mostrar (2016, p. 12).

Pensando no uso formativo das avaliações nas escolas e nas salas de aula, também é importante reconhecer suas limitações:

Os resultados das provas padronizadas não refletem as complexas dinâmicas educacionais que operam no nível das escolas e salas de aula e, por esse motivo, são insuficientes para orientar as práticas pedagógicas (Martínez Rizo, 2016).

É necessário superar a lógica tecnocêntrica do uso de resultados que não reconhecem o processo, as influências e tensões da elaboração de políticas. É necessário considerar a complexidade da mudança educacional e definir o papel que a avaliação pode desempenhar nela: a capacidade de mostrar áreas de melhoria, mas não mostra o caminho a seguir (Ravela, 2015).

Ravela sugere que, em vez de esperar que as mudanças venham da avaliação em si, esta deve servir para confrontar as políticas e indicar onde as mudanças devem ser realizadas:

Precisamos de uma avaliação que não substitua as políticas educacionais, mas as confronte (...) no sentido de sugerir caminhos e promover políticas mais apropriadas. A avaliação deve contribuir para restituir o lugar e a responsabilidade da política educacional e não deve substituí-la (2015, p. 81).

Outro ponto de discordância é a consideração do custo-benefício da aplicação das avaliações:

Como é possível observar, há mais de uma década se realizam avaliações em larga escala que implicam o investimento de uma quantidade significativa de recursos, mas o que ainda não está claro – e é amplamente polêmico – é quão úteis elas têm sido para o sistema educacional nacional e para a sociedade mexicana (Jiménez Moreno, 2016, p. 6).



Posicionamentos mais radicais propõem a eliminação das avaliações e a alocação desse recurso diretamente às ações de melhoria do sistema educacional, indicando o retrocesso ou a estagnação dos resultados como uma prova do fracasso desses sistemas de avaliação<sup>29</sup> (Casassus, 2007). Outros grupos, como na Guatemala, demandam um maior aproveitamento das informações pelo Ministério, propondo seu uso nas políticas de credenciamento de estudantes ou na certificação de centros (Fortín Morales, 2013).

Sobre isso, Ravela et al. (2007) relatam as vantagens, desvantagens e desafios previstos na escolha de sistemas de avaliação de certificação (censos) ou de diagnóstico (geralmente amostras). Uma das questões que consideram é o custo que implica, mas, em todo caso, defendem que as avaliações de qualidade exigem investimento por parte do Estado. Os autores denunciam a existência de governos que demandam avaliações úteis, de alta qualidade e de baixo custo. Além disso, destacam a importância de considerar e equilibrar o investimento da avaliação e a divulgação de seus resultados e a promoção de seu uso.

Por sua vez, Bracho (2018) propõe desenvolver políticas para o uso dos resultados que contemplem diferentes atores desde o momento do desenho das avaliações, para aumentar os usos efetivos dos resultados e, assim, contribuir para elevar a qualidade da educação.

Do ponto de vista teórico do assunto, referindo-se à participação em avaliações internacionais, Addey e Sellar (2019) propõem ampliar o debate sobre o custo-benefício para considerar as motivações de participação (ou não participação) nas provas. A discussão sobre os custos ou benefícios das avaliações incluiria não apenas o uso direto dos resultados, mas também as capacidades técnicas geradas, o pertencimento à comunidade regional/internacional, a legitimidade política e a busca do apoio racional às políticas. Como exemplo, mencionam o caso da Bolívia, que justificou sua não participação no PISA como uma preferência por estabelecer seus próprios referenciais educacionais sem influências externas ou comparações com outros países que não compartilham seu contexto ou ideário de desenvolvimento. Nesse sentido, o custo de dispensar as avaliações se

---

<sup>29</sup> Anteriormente, o documento menciona o aviso de Jiménez Moreno (2016) sobre esse tipo de interpretações, que não consideram outros fatores que influenciam os resultados. O autor apresenta o caso do México, no qual a incorporação de populações anteriormente excluídas do sistema também pode explicar por que não foram observados os avanços nacionais esperados na aprendizagem.

contrapõe ao benefício – na perspectiva do país – de autodeterminar sua agenda educacional.

Finalmente, há uma série de críticas a respeito dos efeitos indesejados que os diferentes sistemas de avaliação produziram, de acordo com a lógica que os guiaram.

Em geral, as autoridades são acusadas de buscar atalhos simplificadores que pretendem acelerar os processos educacionais, implementando políticas que prejudicam o sistema e o aprendizado (Rivas, 2015; Ravela, 2015).

Em primeiro lugar, o que se critica é que os sistemas de avaliação com altas consequências aumentam a desigualdade e prejudicam as escolas em contextos mais desfavorecidos. No caso do Chile, os resultados das avaliações desempenham um papel decisivo na matrícula e nos recursos escolares, uma vez que, em teoria, os pais escolhem as melhores escolas para seus filhos. Com o tempo, foram instaladas práticas de segmentação nas quais as escolas subsidiadas ou particulares escolhem os melhores estudantes e as escolas municipais ficam para trás (Sánchez Cerón e Corte Cruz del Sagrario, 2013). Essa dinâmica de seleção aumenta as desigualdades sociais, uma vez que as escolas de baixos recursos se concentram em aumentar as pontuações em detrimento da aprendizagem, enquanto as escolas particulares pagam cursos preparatórios (Casassus, 2007). A rejeição dessas práticas tem sido categórica por alguns setores da sociedade e, mais recentemente, pela própria ACE. Ayala e Sepúlveda (2018) reconhecem que foram geradas reações contra as avaliações que exigem, inclusive, sua eliminação. Um exemplo disso é o movimento “Alto al SIMCE” (“Parem o SIMCE, em tradução livre), conformado por acadêmicos, professores e estudantes (Sandoval-Hernández, 2016). Sem mencionar casos específicos, o relatório do PREAL desaconselha as práticas de seleção de estudantes para melhorar os resultados das escolas e alerta contra a adoção de expectativas de aprendizagem diferenciadas por escolas de diferentes estratos sociais (Ravela, et al., 2007).

Em 1998, Carol Weiss já identificava, em um plano internacional, que os aspectos avaliados costumam se tornar os aspectos prioritários. Nesse sentido, ainda hoje há uma preocupação com a negligência de outros aspectos educacionais, como as humanidades, os valores e a arte em favor de

disciplinas como linguagem e matemática, promovidas pelas avaliações padronizadas (Sánchez Cerón e Corte Cruz del Sagrario, 2013). Vários autores coincidem que a ênfase excessiva nos resultados das avaliações padronizadas empobrece a educação, gerando uma visão em túnel na qual temas não avaliados e de desenvolvimento pessoal são desprezados, ainda que sejam altamente valorizados pelos próprios professores (Casassus, 2007; Rivas, 2015; Sánchez Cerón e Corte Cruz del Sagrario, 2013).

No plano pedagógico, as avaliações padronizadas são acusadas de homogeneizar as práticas ignorando o contexto dos alunos, e de eliminar a flexibilidade de metodologias e inovação dos professores (Sánchez Cerón e Corte Cruz del Sagrario, 2013). Um exemplo é ensinar os professores a elaborar simulados de avaliações padronizadas para que os estudantes estejam familiarizados com elas (Silva et al., 2012).

Alguns autores sustentam que esses sistemas pervertem a motivação intrínseca dos professores e supõem que precisam de uma coerção externa para trabalhar bem (Casassus, 2007; Sánchez Cerón e Corte Cruz del Sagrario, 2013). Por outro lado, a avaliação do programa piloto de incentivos a docentes na Colômbia mostrou uma melhora nos resultados dos estudantes cujos professores receberam incentivos. Note-se, no entanto, que as práticas realizadas nessas escolas não foram documentadas (Cabal, Milena e Castro, 2018). Ravela (2015) argumenta que, embora o sistema não deva funcionar exclusivamente a partir de exigências externas, a natureza humana exige um certo nível de exigência que vem de fora. Aparentemente, é necessário aprofundar a dinâmica de motivação extrínsecas e intrínsecas, que podem ser mais eficazes para os docentes da região.

Por outro lado, quando se associam incentivos ou altas consequências às escolas ou docentes com base nos resultados de seus estudantes, são geradas práticas de simulação ou teste para o exame. Em uma análise comparativa entre Brasil, México e Chile, são descritas práticas de simulação ou trapaça, que consistem em treinar os estudantes para responder aos exames<sup>30</sup> e solicitar aos menos preparados que não façam a prova (Sánchez Cerón e Corte Cruz del Sagrario, 2013). No caso do PISA, os materiais foram previamente distribuídos para que os professores trabalhassem com os estudantes, com o objetivo de melhorar os resultados (Silva et al., 2012).

<sup>30</sup> No México, até 2013 foram realizadas semanas Pré-EXCALE, nas quais os alunos praticavam com provas similares ao EXCALE.

Associar esse tipo de consequência aos resultados das avaliações [incentivos aos docentes para os resultados dos estudantes] redireciona os próprios fins da educação, uma vez que os atores são estimulados a optar por caminhos de formação que os levem à resolução eficiente das avaliações, em vez de cobrir os objetivos dos planos e programas educacionais (Jiménez Moreno, 2016, p. 16).

Diferentes opiniões foram encontradas sobre o interesse dos estudantes em relação às provas, o que parece depender das possíveis consequências que elas tenham para eles (Fortín Morales, 2013), embora este seja um assunto ainda pouco explorado.

## 7) Conclusões

A partir da literatura revisada neste exercício, pode-se concluir que, embora exista uma disparidade na maturidade dos sistemas de avaliação entre países, há uma tradição de avaliação instalada na região com importantes avanços na transparência dos resultados. O desafio pendente parece ser o uso das informações publicadas para influenciar a melhoria da educação.

Existem marcos de referência sólidos, cada vez mais detalhados, para identificar as abordagens e teorias de mudança que são atribuídas às avaliações em termos de gestão do sistema educacional.

Do ponto de vista empírico, essas estruturas e teorias foram postas à prova, evidenciando algumas de suas limitações a partir das consequências não intencionais que foram geradas, seja por elementos de seu desenho técnico, pelo desenho dos esquemas de consequências, das capacidades para o uso de resultados, da divulgação destes, ou das conjunturas e interesses políticos.

Embora, por um lado, exista uma sistematização bastante concreta das capacidades e limitações técnicas nos países para incluir fatores associados à conquista dentro de modelos, índices ou avaliações mais complexas, que superem as informações descritivas – assim como os órgãos técnicos das áreas operacionais e docentes sobre a compreensão dos resultados das provas e como eles podem usar as informações de acordo com suas funções e âmbitos de ação; também há menos informação sobre as interpretações

e usos precisos que diferentes atores fazem desses resultados além dos exercícios de diagnóstico ou acompanhamento.

O debate sobre a relevância desses sistemas inclui pontos de vista opostos, com uma forte carga crítica, sustentada em grande parte pelas consequências não intencionais das avaliações. No entanto, vozes de convergência emergem em direção a uma concepção mais integral e medida dos sistemas de avaliação como parte articulada dos sistemas educacionais, com um conceito mais amplo de qualidade educacional.

A realização dessa visão integral das avaliações em larga escala requer um conhecimento detalhado das dinâmicas de operação e da inter-relação entre os atores do sistema educacional em todos os níveis. Reduzir a conversa a aspectos técnicos ou comunicativos seria ignorar a natureza complexa e sistêmica do processo de melhoria educacional.

A partir dos trabalhos revisados, surgem as seguintes questões que devem ser abordadas em estudos futuros:

No nível teórico, que esclareça o papel das avaliações e sua articulação dentro dos sistemas educacionais:

- Que fatores mediam as expectativas dos diferentes atores sobre as avaliações em larga escala e sob que lógicas de mudança estão operando?
- Que tipos de avaliações ou estudos devem ser feitos sobre as dinâmicas entre componentes e atores do sistema para superar uma concepção deles como um material estático do processo educacional?
- Que esquemas de motivação interna e de incentivos externos favoreceriam a implementação sistêmica de uma cultura de avaliação para a melhoria no nível escolar e o uso formativo dos resultados?
- Como isso se conjuga com as políticas de prestação de contas?

Em um nível mais operacional e técnico que permita desentranhar as dinâmicas do sistema para atuar sobre elas:

- Como se inter-relacionam nacionalmente os fatores políticos, técnicos e institucionais em relação ao uso observado dos resultados da avaliação?<sup>31</sup>
- Atualmente, qual o papel dos resultados das avaliações nos diferentes

<sup>31</sup> Há informações sobre as capacidades técnicas dos professores, mas não há pesquisas suficientes em nível nacional sobre os posicionamentos, atitudes, relações e capacidades técnicas dos atores técnicos de níveis médios e altos.

momentos de elaboração, implementação e seguimento das políticas e programas para os atores envolvidos?

- Como isso impacta ou explica a diferença de visão e funções esperadas da avaliação pelas partes envolvidas, na elaboração e no uso que é feito delas?
- Existem fatores, além dos identificados, que influenciam o uso observado dos resultados das avaliações?
- Que informações adicionais são requeridas para serem utilizadas com os resultados das avaliações, a fim de se refletirem em políticas, programas e o processo de ensino-aprendizagem?

## Bibliografía

- Addey, C. y Selar, S. (2019). Is it worth it? Rationales for (non)participation in international large-scale learning assessments. Documentos de trabalho, UNESCO, Education Research and Foresight.
- Ayala, C. e Sepúlveda, M. (2018). Vanguardia en evaluación y mejora educativa. La agencia de la Calidad de la Educación en Chile. (R. Red, Entrevistador). Instituto Nacional de Avaliação da Educação. Disponível em: <https://www.inee.edu.mx/index.php/publicaciones-micrositio/blog-revista-red/637-blog-revista-red/blog-revista-red-articulos/3417-vanguardia-pt-avaliacao-e-melhoria-educacional-chile>
- Bracho, T. (2018). Uso de la evaluación: asunto pendiente. (D. i. trabalho, ed.). Instituto Nacional para a Avaliação da Educação.
- Cabal, J., Milena, D. e Castro, F. (2018). Programa de Incentivos Docentes: Impactos y recomendaciones en el corto plazo. Departamento Nacional de Planejamento, Bogotá. Disponível em: [https://colaboracion.dnp.gov.co/CDT/Sinergia/Documentos/Evaluacion\\_incentivos\\_docentes\\_Documento\\_vf.pdf](https://colaboracion.dnp.gov.co/CDT/Sinergia/Documentos/Evaluacion_incentivos_docentes_Documento_vf.pdf)
- Casassus, J. (2007). El precio de la evaluación estandarizada: la pérdida de la calidad y la segmentación social. (ANPAE, ed.) Revista Brasileira de Política e Administração da Educação, 23 (1), p. 71-79.
- CNSP (2019). Lineamientos generales del Sistema de Asesoría y Acompañamiento a las Escuelas. Secretaria de Educação Pública.
- CPEUM (2019). Constituição Política dos Estados Unidos Mexicanos. Diario Oficial da Federação. Disponível em: [https://www.dof.gob.mx/nota\\_detalle.php?codigo=5560457&fecha=15/05/2019](https://www.dof.gob.mx/nota_detalle.php?codigo=5560457&fecha=15/05/2019)
- EDUY21 (sf). La evaluación de aprendizajes en una agenda de transformación educativa. Libro blanco. Montevideu: EDUY21 Iniciativa Cidadã. Recuperado em Recuperado em: <http://www.eduy21.org/Documentos/1-Evaluacion-y-transparencia.pdf>
- FEREMA (2017). Educación: Una dedua pendiente. Informe de progreso educativo Honduras. Tegucigalpa: Fundação para a Educação Ricardo Ernesto Maduro Andreu.
- Ferrer, G. (2006). Educational Assessment Systems in Latin America: Current Practices and Future Challenges. Washington D. C: PREAL.
- Ferrer, G. e Fiszbein, A. (2015). ¿Qué es lo que ha sucedido con los sistemas de evaluación de aprendizajes en América Latina? Lecciones de la última década de experiencia. Grupo Banco Mundial.
- Fortín Morales, Á. M. (2013). Evaluación Educativa Estandarizada en Guatemala: Un camino recorrido, un camino por recorrer. Guatemala: Ministerio da Educação da Guatemala.
- Fundación para la Educación Ricardo Ernesto Maduro Andreu (2017). Informe de progreso educativo Honduras 2017. FEREMA.

- Hanushek, E. y Woessmann, L. (2010). The high cost of low educational performance. The long-run economic impact of improving PISA outcomes. OECD.
- ICFES (2017). Informe nacional de resultados. Colombia en PISA 2015. Instituto Colombiano para a Avaliação da Educação, Bogotá D.C. Consultado em 3 de junho de 2019.
- INEE (2018). Planea: nueva generación de pruebas. ¿Qué evalúa? Disponível em: [https://www.inee.edu.mx/wp-content/uploads/2018/12/Planea\\_3.pdf](https://www.inee.edu.mx/wp-content/uploads/2018/12/Planea_3.pdf)
- INEE (20 de maio de 2019). Instituto Nacional para a Avaliação da Educação. Disponível em: [www.inee.edu.mx](http://www.inee.edu.mx)
- INEVAL (2018). Educación en Ecuador. Resultados de PISA para el desarrollo. Ineval, Quito, Equador. Disponível em: <http://www.evaluacion.gob.ec/evaluaciones/pisa-documentacion/>
- INEVAL (2019). Instituto Nacional de Avaliação Educacional. Disponível em: <https://www.evaluacion.gob.ec/>
- Jiménez Moreno, J. A. (2016). El papel de la evaluación a gran escala como política de rendición de cuentas en el Sistema Educativo Mexicano. Universidade Autônoma de Madri e Red Iberoamericana de Investigación sobre Cambio, Edits.) Revista Iberoamericana de Evaluación Educativa, 9(1), p. 109-126.
- Márquez Jiménez, A. (2017). A 15 años de PISA: resultados y polémicas. (ISSUE-UNAM, Ed.) Perfiles Educativos, XXXIX (156), Editorial.
- Kautz, T., Heckman, J., Diris, R., terWeel, B. e Borghans, L. (2014). Fostering and Measuring Skills: Improving Cognitive and Non-cognitive Skills to Promote Lifetime Success. Paris: OECD Publishing. Disponível em: <https://doi.org/10.1787/5jxsr7vr78f7-en>
- Ley 20529 (2011). Ley del Sistema Nacional del Aseguramiento de la Calidad de la Educación Parvularia, Básica y Media y su Fiscalización. Santiago: Congreso Nacional do Chile. Disponível em: <http://bcn.cl/1uv5c>
- LINEE (2013). Ley del Instituto Nacional para la Evaluación de la Educación. Diário Oficial da Federação. Disponível em: [http://www.dof.gob.mx/nota\\_detalle.php?codigo=5313842&fecha=11/09/2013](http://www.dof.gob.mx/nota_detalle.php?codigo=5313842&fecha=11/09/2013)
- LOEI (2011). Ley Orgánica de Educación Intercultural. Quito: Registro Oficial do Órgão do Gobierno do Equador. Disponível em: [https://oig.cepal.org/sites/default/files/2011\\_leyeducacionintercultural\\_ecu.pdf](https://oig.cepal.org/sites/default/files/2011_leyeducacionintercultural_ecu.pdf)
- Loza, S. E. (2018). Análisis a la evaluación educativa estandarizada en Colombia. Ensayo. Bogotá, Colômbia.
- Martínez Rizo, F. (2008). La evaluación de aprendizajes en América Latina. Distrito Federal, México: INEE. Disponível em: <https://www.inee.edu.mx/wp-content/uploads/2019/01/P1C140.pdf>
- Martínez-Rizo, F. (2016). Impacto de las pruebas de gran escala en contextos de débil tradición técnica: experiencia de México y el Grupo Iberoamericano de PISA. Revista Electrónica de Evaluación Educativa RELIEVE, 22(1), p. 1-12. Disponível em:



- [http://www.uv.es/RELIEVE/v22n1/RELIEVEv22n1\\_M0.pdf](http://www.uv.es/RELIEVE/v22n1/RELIEVEv22n1_M0.pdf)
- Mexicanos Primero (2012). Ahora es cuando. Metas 2012-2024. Cidade do México: Mexicanos Primero Visión 2030, A.C. Disponível em: [http://www.mexicanosprimero.org/images/stories/mp\\_recursos/mp\\_publicaciones\\_de\\_mexicanos\\_primero/AhoraEsCuando2012-2024MetasWEB.pdf](http://www.mexicanosprimero.org/images/stories/mp_recursos/mp_publicaciones_de_mexicanos_primero/AhoraEsCuando2012-2024MetasWEB.pdf)
- Mineducación (2015). Colombia, la mejor educada en el 2025. Líneas estratégicas de la política educativa del Ministerio de Educación Nacional. Colômbia. Disponível em: <https://www.mineducacion.gov.co/1759/w3-article-356137.html>
- Mosqueira, A. e Robinson, D. (2018). Análisis sobre la Evaluación de la Calidad Educativa en America Latina: Caso Colombia. (Universidade Autônoma de Madri e Red Iberoamericana sobre Cambio y Eficacia Escolar, Edits.) Revista Iberoamericana de Evaluación Educativa, 11(1), p. 43-55. Disponível em: <https://doi.org/10.15366/riee2018.11.1.003>
- OECD (2019). Meeting expectations? Reflections on PISA for Development from participating countries. OECD. Disponível em: [https://www.oecd.org/pisa/pisa-for-development/Brief\\_29\\_Meeting\\_expectations.pdf](https://www.oecd.org/pisa/pisa-for-development/Brief_29_Meeting_expectations.pdf)
- Porter, C. (2010). What Shapes the Influence Evidence Has on Policy? The Role of Politics in Research Utilisation. Universidade de Oxford, Department of International Development. Young Lives.
- Prats, J. (2011). Qué explica el éxito mediático del informe PISA. Escuela (3,8999). Disponível em: [http://www.ub.edu/histodidactica/images/documentos/pdf/exito\\_mediatico\\_informe\\_pisa.pdf](http://www.ub.edu/histodidactica/images/documentos/pdf/exito_mediatico_informe_pisa.pdf)
- Ravela, P. (2005). Para comprender las evaluaciones educativas. Fichas didácticas. Montevideo: PREAL. Disponível em: <http://repositorio.minedu.gob.pe/bitstream/handle/123456789/4999/Fichas%20did%20c%20a1cticas%20para%20comprender%20las%20evaluaciones%20educativas.pdf?sequence=1&isAllowed=y>
- Ravela, P. (2015). El uso de los resultados de las evaluaciones para mejorar los procesos de enseñanza. En M. Poggi, Mejorar los aprendizajes en la educación obligatoria. Políticas y Actores, p. 69-112. IIPE UNESCO Buenos Aires.
- Ravela, P., Arregui, P., Valverde, G., Wolfe, R., Ferrer, G., Martínez-Rizo, F. e Wolff, L. (2007). Las evaluaciones educativas que América Latina necesita. Resultados de grupo de trabalho, Programa de Promoção da Reforma Educacional na América Latina e Caribe (PREAL), Grupo de trabalho sobre Padrões e Avaliação, S/D. Disponível em: [https://www.researchgate.net/publication/28216807\\_Las\\_Evaluaciones\\_Educativas\\_que\\_America\\_Latina\\_Necesita](https://www.researchgate.net/publication/28216807_Las_Evaluaciones_Educativas_que_America_Latina_Necesita)
- Rivas, A. (2015). América Latina después de PISA. Lecciones aprendidas de la educación en siete países (2000-2015) (1a. ed.). (F. CIPPEC, Ed.) Cidade Autônoma de Buenos Aires, Argentina.
- Sánchez Cerón, M. e Corte Cruz del Sagrario, F. M. (2013). Las evaluaciones estandarizadas: sus efectos en tres países lationamericanos. Revista Lationamericana de Estudios Educativos, XLIII(1), p. 97-124.

- Sánchez, H. (2018). Autonomía de los organismos de evaluación: ¿voluntad o derecho? (INAP, Ed.) Estado abierto. Revista sobre el Estado, la administración y las políticas públicas, 2(3), p. 15-45. Disponível em: <https://tecnoadministracionpub.files.wordpress.com/2018/11/estado-abierto-vol-2-n-3.pdf>
- Sandoval-Hernández, A. (2016). Experiencias internacionales sobre el uso de los resultados de evaluaciones educativas. México: INEE.
- Secretaría de Evaluación Educativa. (2017). Aprender 2017. Informe de resultados Primaria. Ministério da Educação da Argentina.
- Silva, M., D'Angelo, N., Fernández, L., Martínez, O. y Rodríguez, A. (2012). El uso de los resultados de las evaluaciones del INEE por las autoridades educativas. Cidade do México: INEE.
- UNESCO (2014). A brief history of LLECE. UNESCO Office in Santiago Newsletter (14º). Oficina de UNESCO en Santiago. Disponível em: <http://www.unesco.org/new/en/santiago/press-room/newsletters/newsletter-laboratory-for-assessment-of-the-quality-of-education-llece/n14/a-brief-history-of-llece/>
- Urquiola, M. (2011). Calidad y cantidad educativa en Bolivia: 1996-2010. Informe preparado para el Ministerio de Relaciones Exteriores de Holanda. Minister van Buitenlandse Zaken, Policy and Operations Evaluation Department (IOB), El Haya. Disponível em: <http://archieff.iob-evaluatie.nl/sites/iob-evaluatie.nl/files/351%20subrapport%20Urquiola%20Bolivia.pdf>
- Urquiola, M. (2011). Calidad y cantidad educativa en Bolivia: 1996-2010. Informe preparado para el Ministerio de Relaciones Exteriores de Holanda. Bolívia. Disponível em: [http://www.iob-evaluatie.nl/sites/iob-evaluatie.nl/files/351\\_subrapport\\_Urquiola\\_Bolivia.pdf](http://www.iob-evaluatie.nl/sites/iob-evaluatie.nl/files/351_subrapport_Urquiola_Bolivia.pdf)
- Weiss, C. (1979). The many meanings of research utilization. Public Administration Review, 39(5), p. 426-431. doi:10.2307/3109916
- Weiss, C. (1998). Have we learned anything new about the use of evaluation? American Journal of Evaluation, 19(1), p. 21-33. Disponível em: <https://pdfs.semanticscholar.org/976f/28dfe161d086b2605af6c3e364a7bf8bd991.pdf>



Organização  
das Nações Unidas  
para a Educação,  
a Ciência e a Cultura



Instituto Internacional de  
Planejamento Educacional  
IIPE UNESCO Buenos Aires  
Escritório para a América Latina

Planejar **a educação**, construir  **FUTURO**