

NOTA TÉCNICA N° IDB-TN- 03003

El estado de la educación en América Latina y el Caribe 2024: la medición de los aprendizajes

Elena Arias Ortiz
Ximena Dueñas
Cecilia Giambruno
Ángela López

Banco Interamericano de Desarrollo
División de Educación

Setiembre 2024



El estado de la educación en América Latina y el Caribe 2024: la medición de los aprendizajes

Elena Arias Ortiz
Ximena Dueñas
Cecilia Giambruno
Ángela López.

Banco Interamericano de Desarrollo
División de Educación

Setiembre 2024



Catalogación en la fuente proporcionada por la
Biblioteca Felipe Herrera del
Banco Interamericano de Desarrollo
El estado de la educación en América Latina y el Caribe 2024: la medición de
los aprendizajes / Elena Arias Ortiz, Ximena Dueñas, Cecilia Giambruno,
Ángela López.

p. cm. — (Nota técnica del BID ; 3003)

Incluye referencias bibliográficas.

1. Education-Evaluation-Latin America. 2. Education-Evaluation-Caribbean
Area. 3. Education-Research-Latin America. 4. Education-Research-
Caribbean Area. 5. Education and state-Latin America. 6. Education and
state-Caribbean Area. I. Arias Ortiz, Elena. II. Dueñas, Ximena. III. Giambruno,
Cecilia. IV. López Sánchez, Ángela. V. Banco Interamericano de Desarrollo.
División de Educación. VI. Serie.
IDB-TN-3003

Códigos JEL: I21; I24

Palabras claves: Educación, Aprendizaje, Evaluaciones de aprendizaje, Calidad
educativa, Equidad educativa, Indicadores de Educación, Políticas educativas,
Evaluaciones a gran escala

<http://www.iadb.org>

Copyright © 2024 Banco Interamericano de Desarrollo (BID). Esta obra se encuentra sujeta a una
licencia Creative Commons CC BY 3.0 IGO (<https://creativecommons.org/licenses/by/3.0/igo/legalcode>).
Se deberá cumplir los términos y condiciones señalados en el enlace URL y otorgar el respectivo
reconocimiento al BID.

En alcance a la sección 8 de la licencia indicada, cualquier mediación relacionada con disputas que
surjan bajo esta licencia será llevada a cabo de conformidad con el Reglamento de Mediación de la
OMPI. Cualquier disputa relacionada con el uso de las obras del BID que no pueda resolverse
amistosamente se someterá a arbitraje de conformidad con las reglas de la Comisión de las Naciones
Unidas para el Derecho Mercantil (CNUDMI). El uso del nombre del BID para cualquier fin distinto al
reconocimiento respectivo y el uso del logotipo del BID, no están autorizados por esta licencia y
requieren de un acuerdo de licencia adicional.

Note que el enlace URL incluye términos y condiciones que forman parte integral de esta licencia.

Las opiniones expresadas en esta obra son exclusivamente de los autores y no necesariamente
reflejan el punto de vista del BID, de su Directorio Ejecutivo ni de los países que representa.



2024

EL ESTADO DE LA **EDUCACIÓN**

en América Latina y el Caribe

LA MEDICIÓN DE LOS APRENDIZAJES

Elena Arias Ortiz, Ximena Dueñas, Cecilia Giambruno y Angela López



EL ESTADO DE LA EDUCACIÓN

en América Latina y el Caribe

2024

La segunda edición del informe "El Estado de la Educación en América Latina y el Caribe 2024" examina la evolución de las evaluaciones de aprendizaje a gran escala en la región. El informe comienza con un panorama educativo a partir de una selección de indicadores sobre recursos, cobertura, eficiencia y aprendizajes, con un enfoque particular en las inequidades socioeconómicas. Luego, se centra en la evolución de la medición de aprendizajes en 18 países desde la década de 1990 hasta la actualidad, destacando hitos importantes en evaluaciones regionales e internacionales y en el diseño e implementación de evaluaciones nacionales a gran escala. Se analizan aspectos como la frecuencia y cobertura de las pruebas, los cambios institucionales, los tipos de evaluaciones y el impacto de la pandemia, así como los desafíos que afectan la calidad y validez de las mediciones. El informe también examina cómo estas evaluaciones influyen en la formulación de políticas educativas y su papel en la mejora de la calidad educativa. Finalmente, ofrece recomendaciones para fortalecer los sistemas de evaluación, apoyar la recuperación del aprendizaje postpandemia y promover una mayor equidad educativa.

CONTENIDO

Introducción	5
Parte I. Panorama de la educación en América Latina y el Caribe	8
I.1 Recursos financieros	8
I.2 Recursos físicos: acceso y uso de tecnología	11
I.3 Cobertura	16
I.4 Eficiencia	19
I.5 Aprendizaje	23
Parte II. Evolución y perspectivas de la medición de aprendizajes	25
II.1 Evaluaciones internacionales y regionales	27
Evaluaciones regionales	28
Evaluaciones internacionales	31
¿Por qué son importantes estas evaluaciones para los sistemas nacionales de evaluación?	33
II.2 Las evaluaciones de aprendizaje nacionales en la región: evolución y estado actual	38
Creación de capacidades evaluativas (1990-2000)	38
Cambios institucionales y ampliación de los sistemas de evaluación educativa (2001-2010)	42
Consolidación de los sistemas nacionales de evaluación (2011-2019)	47
Disrupción y estado actual de las pruebas (2020-2023)	52
II.3 Reflexiones finales: hacia una cultura de evaluación educativa	60
Anexo	63
Referencias	64

Índice de figuras

Figura 1: gasto en educación como porcentaje del PIB (Circa 2023)	9
Figura 2: financiamiento inicial de gobierno por estudiante de primaria en dólares PPA (Circa 2022)	10
Figura 3: financiamiento inicial de gobierno por estudiante de secundaria en dólares PPA (Circa 2022)	11
Figura 4: acceso a recursos digitales en ALC y en la OCDE (2022)	12
Figura 5: acceso a computadoras por país (2022)	13
Figura 6: acceso a internet por país (2022)	13
Figura 7: acceso a computadora en la institución educativa por nivel socioeconómico (2022)	14
Figura 8: acceso a internet en la institución educativa por nivel socioeconómico (2022)	14
Figura 9: uso de dispositivos digitales por materia (2022)	16
Figura 10: brecha en la cobertura neta de educación primaria por nivel socioeconómico (Circa 2023)	17
Figura 11: brecha en la cobertura neta de educación secundaria por nivel socioeconómico (Circa 2023)	18
Figura 12: brecha en la cobertura neta de educación terciaria por nivel socioeconómico (Circa 2023)	19
Figura 13: tasa de abandono temprano por nivel socioeconómico (Circa 2023)	20
Figura 14: tasa de terminación de educación secundaria por nivel socioeconómico (Circa 2023)	21
Figura 15: tasa de terminación de educación terciaria por nivel socioeconómico (Circa 2023)	22
Figura 16: rezago en años de escolaridad en comparación al promedio OCDE (2022)	24
Figura 17: cantidad de países participantes en ERCE	29
Figura 18: cantidad de países participantes en PISA	32
Figura 19: países con pruebas de aprendizaje a gran a escala por nivel educativo, 18 países (1990-2023)	39
Figura 20: grados evaluados en pruebas de aprendizaje a gran a escala por nivel educativo, 18 países (1990-2023)	41
Figura 21 . países con pruebas nacionales de aprendizaje a gran a escala por tipo de prueba y nivel educativo, 18 países (1990-2023)	43

Índice de tablas

Tabla 1: participación de países de ALC en pruebas internacionales IEA	33
Tabla 2: aplicación de evaluación(es) por año, 1990 – 2023 18 países	54
Tabla 3: tipo de prueba e implicancias antes, durante y luego de la disrupción de la pandemia	56
Tabla 4: presencia de evaluación por año y tipo de institución evaluadora, 2000 – 2023 18 países	57

INTRODUCCIÓN

La medición de los resultados de aprendizaje es fundamental para mejorar la calidad de la educación. Al evaluar el aprendizaje de sus estudiantes, los y las educadores pueden identificar lagunas en la comprensión y áreas que necesitan mejoras. Este enfoque basado en datos permite realizar intervenciones específicas y apoya la mejora continua tanto de las prácticas de enseñanza como del rendimiento de los y las estudiantes. Podemos resumir la intencionalidad de la evaluación en dos frases atribuidas al físico y matemático británico William Thomson Kelvin: “lo que no se define no se puede medir; y lo que no se mide no se puede mejorar.” La evaluación no es el fin: es el medio que permite enfocarse en el uso de resultados para mejorar y, así, ayuda a que los sistemas educativos desarrollen las habilidades y el conocimiento que los y las estudiantes necesitan para tener éxito en el mundo laboral, vivir en sociedad y desarrollar vidas plenas.

Es difícil estar en desacuerdo con estas ideas, pero su veracidad depende de un supuesto implícito: que contamos con las herramientas adecuadas para la medición de los aprendizajes. Para saber en qué medida las intervenciones, programas o políticas educativas inciden en la mejora de los aprendizajes, debemos asegurar que cada actor –estudiantes, docentes, directores, supervisores y responsables de política– cuente con las herramientas de medición adecuadas para determinar si estamos avanzando en la dirección correcta. Es clave contar con sistemas nacionales de evaluación educativa sólidos que ayuden a identificar aquellas escuelas y estudiantes que tienen más desafíos y orientar la inversión de los recursos hacia aquellas que más los necesitan. En los países de América Latina y el Caribe, las evaluaciones de aprendizaje fijan claramente este punto como norte de la política educa-

tiva: mejorar el nivel de aprendizaje y promover una mayor equidad en los aprendizajes. Los resultados de PISA 2022 (por las siglas en inglés de Programa Internacional para la Evaluación de Estudiantes) subrayan la crisis de aprendizaje en América Latina y el Caribe: uno de cada dos estudiantes (55%) sin habilidades básicas de lectura y tres de cada cuatro (75%) sin competencias básicas en matemáticas. La situación es aún más crítica entre las y los estudiantes más pobres, entre quienes nueve de cada diez (88%) no tienen las habilidades básicas en matemáticas (Arias Ortiz et al., 2023a).

Las evaluaciones a gran escala juegan un papel preponderante en los sistemas educativos alrededor del mundo y son utilizadas por organismos internacionales (OCDE y UNESCO), gobiernos nacionales, estados y municipalidades para medir e informar logros de aprendizajes de un sistema educativo. Las evaluaciones de aprendizaje a gran escala son evaluaciones que miden los resultados de aprendizaje de un grupo específico de estudiantes, según su edad o nivel de curso, en un año académico determinado. Estas evaluaciones se distinguen por su uniformidad y estandarización en cuanto al contenido, el proceso de administración, las fechas y los sistemas de puntuación (UNESCO, 2019). Generalmente de carácter sumativo¹, las evaluaciones de este tipo proporcionan un resumen general del rendimiento académico en áreas específicas y se implementan tanto a nivel nacional como transnacional (regional o internacional). Ciertamente, estas evaluaciones presentan algunas desventajas dado que miden una pequeña parte de lo que las y los estudiantes aprenden en la escuela, generalmente evaluando sólo matemáticas, lenguaje y ciencias, excluyendo otras áreas temáticas (Vegas y Petrow, 2008). Los conocimientos y comportamientos no académicos, como las

¹ Las evaluaciones sumativas son un tipo de evaluación del aprendizaje que se realiza al final de una unidad o secuencia de unidades didácticas para medir el alcance que el educando ha logrado de los resultados esperados (Bernard, 2009).

habilidades para la vida, los valores éticos y morales, las habilidades artísticas y creativas y el sentido de responsabilidad cívica o social son más difíciles de cuantificar y, por lo tanto, a menudo quedan fuera del alcance de este tipo de evaluaciones. Incluso el conocimiento “objetivo”, como los hechos y las habilidades de razonamiento básico, puede variar entre sistemas culturales o de valores y, por lo tanto, ser difícil de evaluar.

A pesar de estas limitaciones, las evaluaciones a gran escala ofrecen muchas ventajas como herramienta de monitoreo en términos de objetividad y generación de datos para la toma de decisiones a nivel del sistema y la formulación de la política educativa en ejes claves como currículo, asignación de recursos y prácticas pedagógicas. Al mismo tiempo, la comparabilidad de los resultados permite mejorar la equidad. Al permitir poner en relación directa el contexto, los recursos y el desempeño, se pueden identificar las brechas de equidad en el aprendizaje y promover políticas que fomenten una distribución más equitativa de las oportunidades educativas. Lamentablemente, la realidad es más compleja, ya que la diversidad de las experiencias de aprendizaje y los distintos contextos en los que los y las estudiantes aprenden –incluyendo variaciones en antecedentes socioeconómicos, influencias culturales y características escolares– hacen que el desarrollo de herramientas de evaluación universales que midan de manera justa el progreso de cada estudiante sea un desafío considerable.

Esta segunda entrega de *El estado de la educación en América Latina y el Caribe* presenta un análisis detallado sobre la situación actual de las evaluaciones de aprendizaje a gran escala en la región, el papel que desempeñan en la toma de decisiones en la política educativa y los retos y las principales oportunidades que ofrecen para mejorar la calidad educativa.

La primera parte del informe presenta, como en cada edición de *El estado de la educación*, una actualización integral del panorama edu-

cativo en la región, basada en indicadores del portal de estadísticas educativas CIMA (Centro de Información para la Mejora de los Aprendizajes). Se analiza el estado de los recursos financieros y físicos disponibles en cada país, la cobertura educativa, la eficiencia interna y los logros de aprendizaje de las y los estudiantes. En esta entrega hacemos particular énfasis en las desigualdades tanto entre países como dentro de ellos, con un análisis detallado de las brechas según el nivel socioeconómico. Los resultados muestran una considerable disparidad en la inversión por estudiante en comparación con el promedio de la OCDE. También se destacan notables diferencias en el acceso a recursos digitales entre los países de la región, lo que impacta especialmente a las y los estudiantes de contextos socioeconómicos más desfavorecidos. La cobertura educativa sigue siendo limitada en los niveles secundario y terciario, y la baja eficiencia de los sistemas se refleja en altas tasas de deserción escolar y bajas tasas de finalización, particularmente en los sectores más vulnerables. Finalmente, los datos de PISA 2022 evidencian las persistentes brechas en equidad de aprendizajes, tanto al interior de los países como en comparación con la OCDE.

La segunda parte del informe se centra en un tema específico: la evolución y el estado actual de los sistemas de evaluación educativa en América Latina y el Caribe. Primero, analizamos la participación de los países en evaluaciones regionales e internacionales, su relevancia en las políticas educativas nacionales y su papel en el seguimiento de las metas del Objetivo de Desarrollo Sostenible 4. Luego analizamos las grandes etapas de evolución de las evaluaciones de aprendizaje nacionales de gran escala² usando una metodología innovadora que incluye un análisis cuantitativo de las evaluaciones realizadas, los niveles educativos y su alcance desde 1990 hasta 2023, con un foco especial en dieciocho países de América Latina (Argentina, Bolivia, Brasil, Chile, Colombia, Costa Rica, Ecuador, El Salvador, Guatemala, Honduras, México, Nicaragua, Panamá, Paraguay, Perú, República Dominicana,

2 El análisis se centró en las evaluaciones nacionales de aprendizaje a gran escala sumativas aplicadas entre 1990 y 2023 durante los ciclos lectivos correspondientes a los niveles de educación primaria y secundaria.

Uruguay y Venezuela). Analizamos la cobertura de la población objetivo (muestra o censo), las áreas y niveles evaluados, la orientación y su frecuencia de administración. También discutimos el impacto de la pandemia en la continuidad de las evaluaciones, considerando las características de los sistemas de evaluación y los tipos de evaluaciones en los países. Finalmente, la última sección sintetiza los principales hallazgos y discute lecciones aprendidas y retos en el camino hacia una verdadera cultura de la evaluación.

Los resultados del análisis revelan una evolución diversa, pero con desafíos comunes. Primero, si bien el monitoreo de los aprendizajes sigue siendo clave en las políticas educativas y la mayoría cuenta con evaluaciones de aprendizaje nacionales para este efecto, el nivel de madurez de los sistemas de evaluación varía considerablemente de país en país. Las evaluaciones transnacionales, como ERCE y PISA, juegan un papel crucial justamente para la formación de capacidades técnicas y para la consolidación de los sistemas de evaluación. Sin embargo, la falta de centros especializados en medición educativa y la escasez de expertos en psicometría son obstáculos que limitan el desarrollo de sistemas de evaluación sólidos en la región. Faltan instancias de validación y transparencia sobre la calidad y validez de las evaluaciones, aspectos críticos para generar confianza en los sistemas de evaluación.

Segundo, la eficiencia en la entrega y uso de resultados sigue siendo un desafío significativo, al igual que el equilibrio entre diferentes tipos de evaluaciones que, como decíamos, sigue muy enfocada en evaluaciones a gran escala. Tercero, las evaluaciones nacionales requieren financiamiento constante pero su continuidad está amenazada por la reducción de recursos. Especialmente para la innovación en términos de metodologías y de áreas evaluadas, como la medición de componentes socioemocionales y el uso de tecnología en las evaluaciones –estas son áreas en crecimiento, aunque la limitada infraestructura

dificulta su implementación a gran escala–. A pesar de los retos, es fundamental retomar y fortalecer los sistemas de evaluación en el contexto postpandemia para guiar la recuperación de los aprendizajes en la región.

El fortalecimiento de los sistemas de evaluación no se limita a las evaluaciones de aprendizaje a gran escala de carácter sumativo, sino que también requiere la integración de otros tipos de mediciones que van más allá del foco de este informe. Entre ellas se encuentran las evaluaciones formativas, diseñadas para ofrecer retroalimentación continua durante el proceso de aprendizaje y ajustar la enseñanza según las necesidades de los alumnos³, las evaluaciones diagnósticas, que con un enfoque formativo permiten evaluar conocimientos y habilidades previas al inicio del año lectivo para identificar áreas de necesidad o potencial de enriquecimiento; y las autoevaluaciones, que fomentan la reflexión de los estudiantes. Cada tipo de evaluación cumple un propósito específico y, en conjunto, proporcionan la retroalimentación necesaria para mejorar continuamente las prácticas de enseñanza. El objetivo final, como mencionamos al principio, es mejorar la educación para aumentar las oportunidades para todos en la región.

³ Mientras la evaluación sumativa se centra en el final de una etapa educativa, la evaluación formativa se lleva a cabo durante toda la etapa, enfocándose en el proceso y permitiendo iteraciones continuas entre estudiante y docente, lo que proporciona información personalizada y detallada para ajustar las estrategias de aprendizaje y enseñanza (UNESCO, 2021).

PARTE I.

Panorama de la educación en América Latina y el Caribe

Un sistema educativo puede conceptualizarse como un proceso en el que se aportan insumos financieros y físicos para facilitar el acceso, la permanencia y el aprendizaje de los y las estudiantes en un país⁴. Para entender mejor esos procesos es necesario tener información. En esa línea, en 2017 el BID creó el Centro de Información para la Mejora de Aprendizajes (**CIMA**) para mejorar la disponibilidad de datos e indicadores comparables en la región. El CIMA proporciona datos que miden cada etapa del proceso educativo y permiten desagregar la información por múltiples dimensiones, incluyendo nivel socioeconómico, género y ubicación geográfica.

En este capítulo presentamos un análisis detallado de la educación en América Latina y el Caribe a través de una serie de indicadores, haciendo especial énfasis en la persistente inequidad entre niveles socioeconómicos en los países de la región. En primer lugar, examinamos los insumos financieros disponibles, utilizando datos de UNESCO⁵. Luego evaluamos el acceso a recursos físicos, con un énfasis en el acceso y uso de recursos digitales, basándonos en datos de la prueba PISA 2022. En la tercera sección analizamos la cobertura educativa en la región mediante tasas de asistencia neta, obtenidas de encuestas nacionales de hogares, con datos de 2021 a 2023. Posteriormente evaluamos la eficiencia del sistema educativo utilizando tasas de abandono y de finalización, también derivadas de las encuestas nacionales de hogares. Finalmente, medimos el aprendizaje analizando los niveles de bajo desempeño y el rezago en años de educación en comparación con el promedio de la OCDE, utilizando nuevamente datos de la prueba PISA.

I.1 Recursos financieros

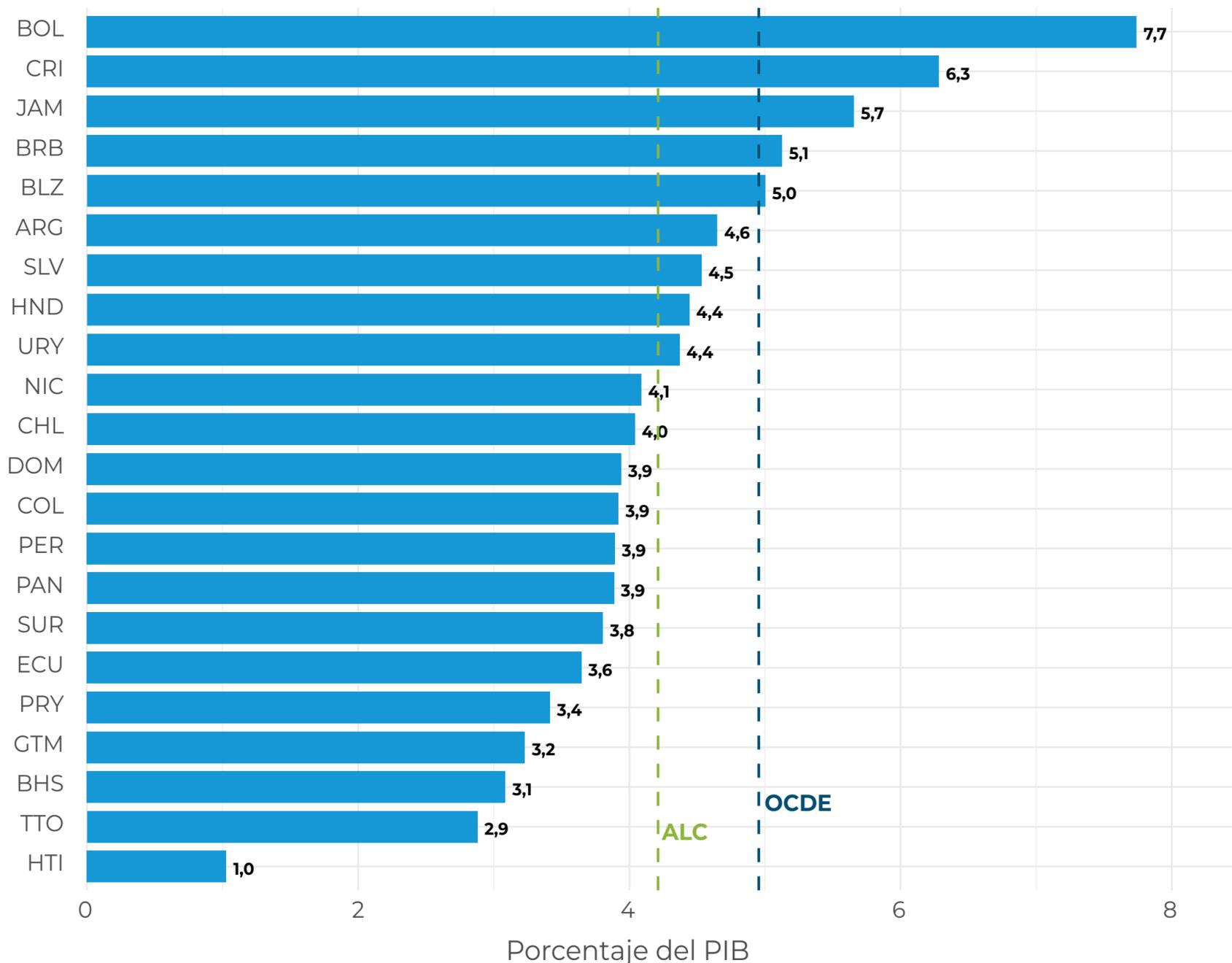
Para analizar los recursos financieros asignados a la educación utilizamos como indicadores el gasto en educación como porcentaje del PIB y el financiamiento inicial del gobierno en dólares constantes en términos de paridad de poder adquisitivo (dólares PPA) por nivel educativo.

El gasto público destinado a la educación como porcentaje del PIB refleja el esfuerzo que un estado dedica a la educación en relación con la capacidad económica nacional (UNESCO Institute for Statistics, 2022). La comparación del financiamiento inicial del gobierno por estudiante en términos de dólares PPA permite examinar la inversión por estudiante en un nivel educativo específico. Este indicador complementa la métrica del gasto público en educación como porcentaje del PIB, dado que un mayor porcentaje no necesariamente se traduce en más recursos disponibles por estudiante (Arias Ortiz, et al., 2023c).

En América Latina y el Caribe, el gasto promedio en educación equivale a 4,2 por ciento del PIB, un nivel apenas inferior al promedio de los países de la OCDE (5 por ciento). Sin embargo, la mitad de los países de la región no alcanza el nivel recomendado por la UNESCO, que sugiere destinar al menos entre 4 y 6 por ciento del PIB a la educación (UNESCO Institute for Statistics, 2022). Efectivamente, existe una considerable disparidad entre los países de América Latina y el Caribe. Los países que destinan más gasto público a la educación como porcentaje del PIB son Bolivia (7,7 por ciento), Costa Rica (6,3 por ciento) y Jamaica (5,7 por ciento); mientras que los países que asignan el menor porcentaje son Haití (1 por ciento), Trinidad y Tobago (2,9 por ciento) y Bahamas (3,1 por ciento) (Figura 1).

⁴ Este capítulo contó con la valiosa contribución de Melchor de la Cruz, pasante de la División de Educación en el marco del Summer Internship Program del BID 2024

⁵ Los datos de Recursos Financieros presentados en este documento provienen de la actualización de febrero de 2024 del Instituto de Estadística de la UNESCO (UIS). Se incluyen indicadores tanto de la colección 'SDG Global and Thematic Indicators' como de la colección 'Other Policy Relevant Indicators'

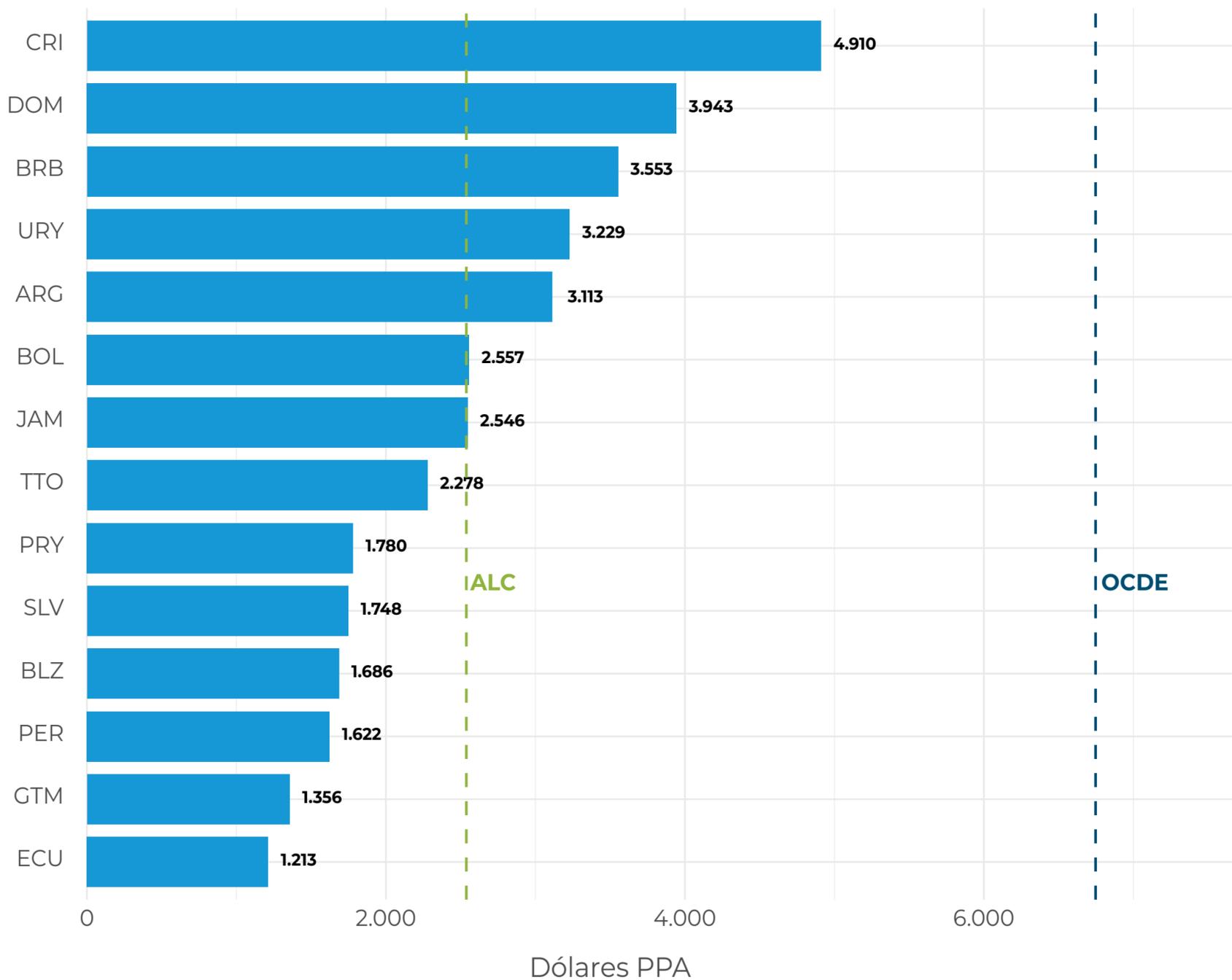
Figura 1: gasto en educación como porcentaje del PIB (Circa 2023)⁶

Fuente: elaboración propia en base a CIMA.

Sin embargo, medida en dólares PPA, hay una brecha significativa en la inversión pública por estudiante entre el promedio de la región y el promedio de la OCDE en la educación primaria. En la OCDE la inversión promedio por estudiante del nivel primario es de 6.700 dólares PPA, mientras que en América Latina y el Caribe es de sólo 2.500 dólares PPA por estudiante, lo que representa 63 por ciento menos.

Todos los países de la región evidencian una inversión por estudiante de primaria menor a la del promedio de la OCDE. Además, existe una notable disparidad dentro de la región. El mayor financiamiento inicial de primaria se encuentra en Costa Rica (4.910 dólares PPA), República Dominicana (3.943 dólares PPA) y Barbados (3.553 dólares PPA). En contraste, Ecuador (1.213 dólares PPA), Guatemala (1.353 dólares PPA) y Perú (1.622 dólares PPA) exhiben el menor financiamiento (Figura 2).

⁶ Los datos de Chile (CHL), Nicaragua (NIC), Surinam (SUR), Colombia (COL), Costa Rica (CRI) y Argentina (ARG) corresponden al año 2021. Los datos de Haití (HTI), Bolivia (BOL), Ecuador (ECU), Perú (PER), República Dominicana (DOM), Panamá (PAN), Paraguay (PRY), El Salvador (SLV), Trinidad y Tobago (TTO), Honduras (HND), Guatemala (GTM) y Uruguay (URY) corresponden al año 2022. Los datos de Belice (BLZ), Jamaica (JAM), Barbados (BRB), Bahamas (BHS) y la OCDE corresponden al año 2023. El valor para América Latina y el Caribe es calculado como promedio simple de los datos de los países.

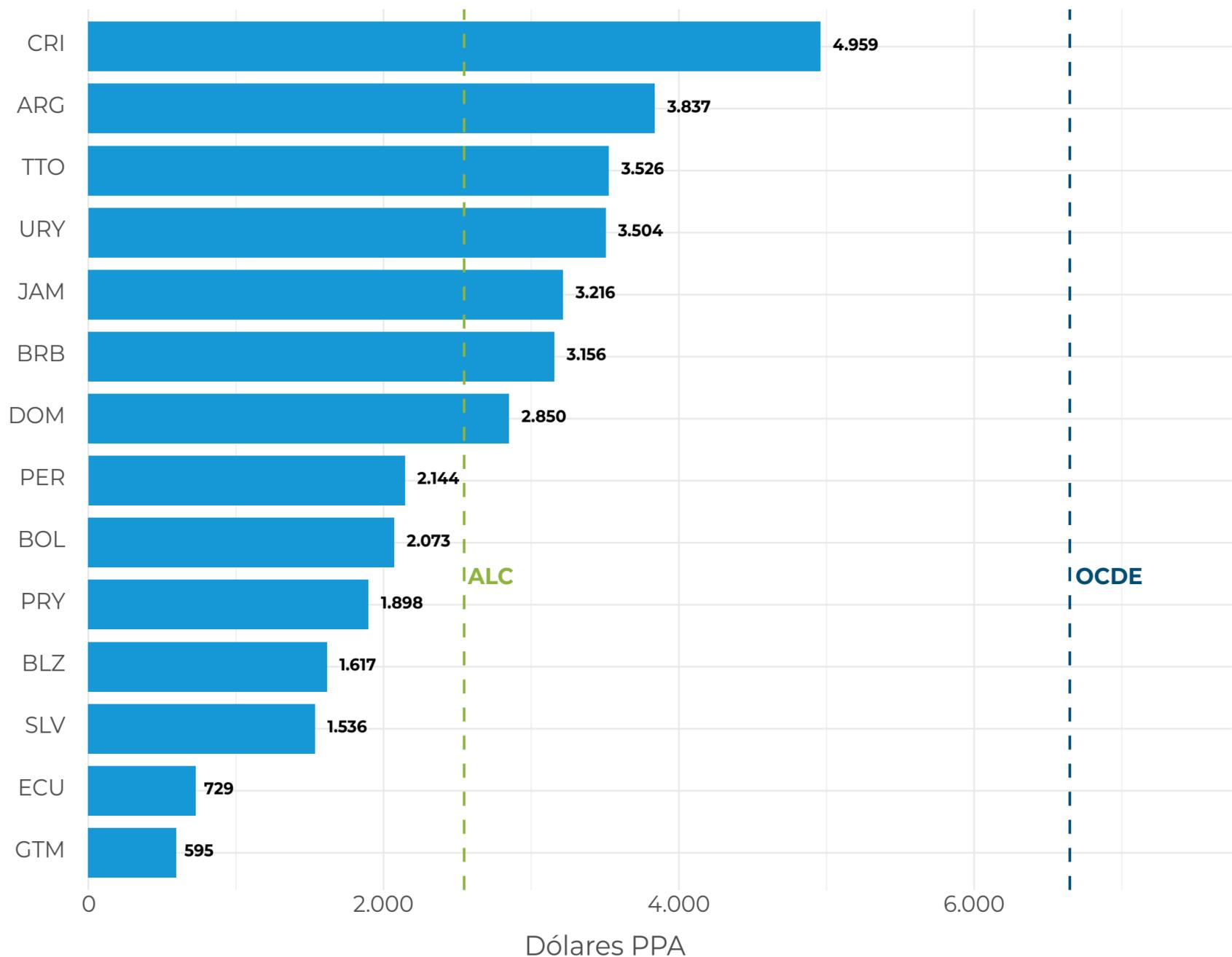
Figura 2: financiamiento inicial de gobierno por estudiante de primaria en dólares PPA (Circa 2022)⁷

Fuente: elaboración propia en base a CIMA.

En América Latina y el Caribe la inversión pública promedio en educación secundaria es de 2.600 dólares PPA, lo que representa 62 por ciento menos que el promedio de la OCDE. Asimismo, hay una considerable disparidad entre los países de la región. El mayor financiamiento inicial de secundaria se encuentra en Costa Rica (4.959 dólares PPA), Argentina

(3.837 dólares PPA) y Trinidad y Tobago (3.526 dólares PPA). En contraste, la inversión por estudiante es menor en Guatemala (595 dólares PPA), Ecuador (729 dólares PPA) y El Salvador (1.536 dólares PPA) (Figura 3).

⁷ Los datos de Trinidad y Tobago (TTO), El Salvador (SLV), Costa Rica (CRI), Uruguay (URY), Argentina (ARG) y la OCDE corresponden al año 2021. Los datos de Bolivia (BOL), Ecuador (ECU), República Dominicana (DOM), Paraguay (PRY), Guatemala (GTM), Belice (BLZ), Perú (PER), Jamaica (JAM) y Barbados (BRB) corresponden al año 2022. El valor para América Latina y el Caribe (ALC) es calculado como promedio simple de los datos de los países.

Figura 3: financiamiento inicial de gobierno por estudiante de secundaria en dólares PPA (Circa 2022)⁸

Fuente: elaboración propia en base a CIMA.

1.2 Recursos físicos: acceso y uso de tecnología

La transformación digital en los sistemas educativos puede ampliar el acceso a oportunidades educativas, promover la inclusión y mejorar la pertinencia y calidad del aprendizaje. Para que esta transformación sea viable es esencial contar con un acceso completo y equitativo a la infraestructura digital. Las pruebas PISA 2022 proporcionan una base valiosa para analizar tanto el acceso a dispositivos digitales como la conectividad en las instituciones educativas.⁹

Además del acceso, los formularios aplicados a las y los estudiantes permiten examinar el uso de dispositivos digitales en los centros educativos de América Latina y el Caribe¹⁰.

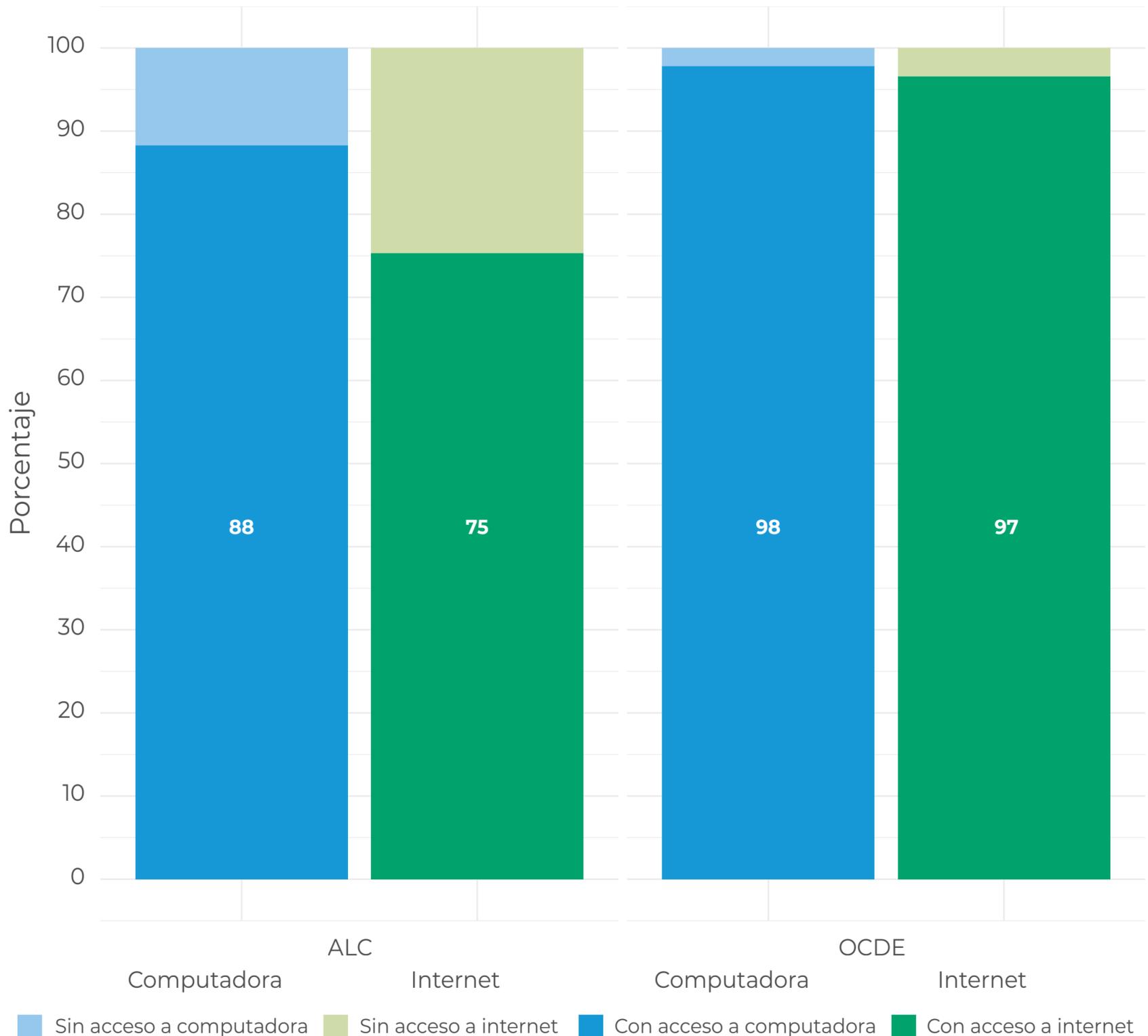
La combinación de estos indicadores facilita la generación de un panorama más completo, lo cual es esencial para el desarrollo de estrategias efectivas de transformación digital en la región.

⁸ Los datos de Trinidad y Tobago (TTO), El Salvador (SLV), Costa Rica (CRI), Uruguay (URY), Argentina (ARG) y la OCDE corresponden al año 2021. Los datos de Bolivia (BOL), Ecuador (ECU), República Dominicana (DOM), Paraguay (PRY), Guatemala (GTM), Belice (BLZ), Perú (PER), Jamaica (JAM) y Barbados (BRB) corresponden al año 2022. El valor para América Latina y el Caribe (ALC) es calculado como promedio simple de los datos de los países.

⁹ Se considera que un o una estudiante tiene acceso a una computadora en la escuela si, en el formulario de directores, se indica que la escuela cuenta con al menos una computadora disponible para el uso de las y los estudiantes. Asimismo, se considera que un o una estudiante tiene acceso a internet en la escuela si, en el mismo formulario, se declara que la escuela dispone de al menos una computadora con acceso a internet disponible para el uso de los y las estudiantes.

¹⁰ Se considera que en una materia se utiliza infraestructura digital cuando se emplean dispositivos digitales en la mitad o más de las clases.

Figura 4: acceso a recursos digitales en ALC y en la OCDE (2022)

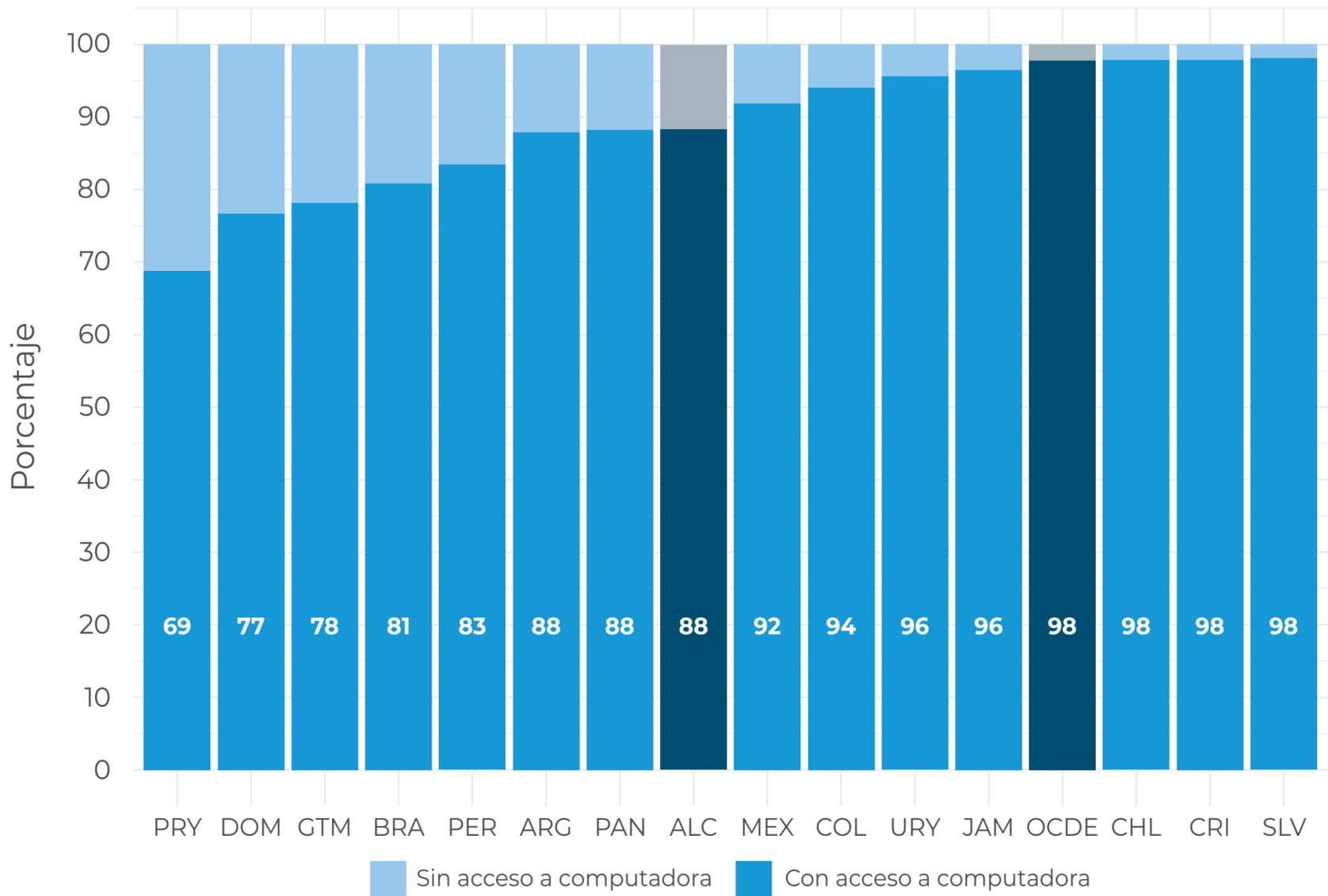


Fuente: elaboración propia en base a CIMA.

En América Latina y el Caribe 1 de cada 10 estudiantes no tiene acceso a una computadora en su institución educativa. De manera similar, 2 de cada 10 estudiantes no tienen acceso a internet en las mismas instituciones. Esto contrasta con el promedio de la OCDE, donde se reporta un acceso casi universal tanto a computadoras como a internet (Figura 4).

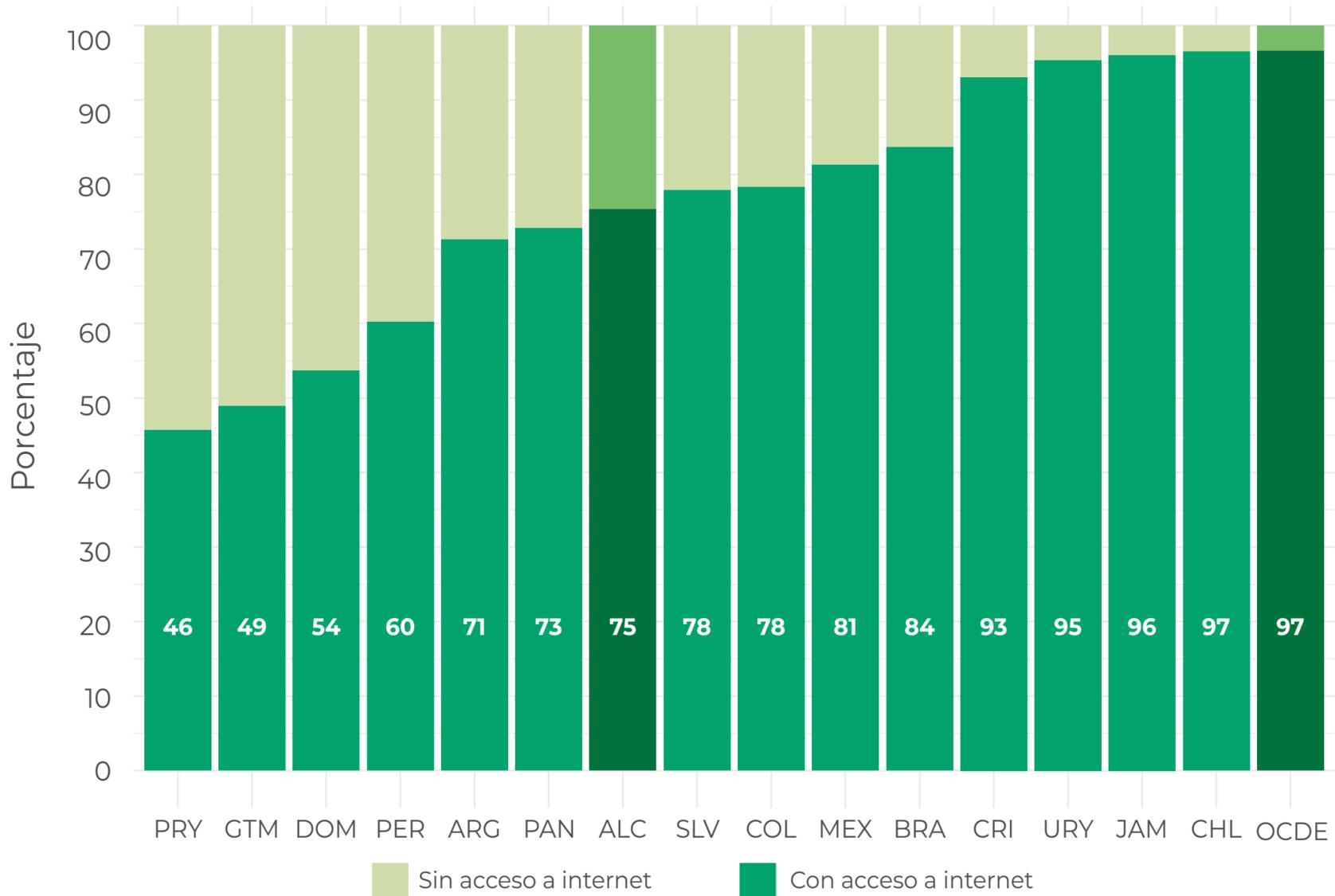
Al comparar a los países de la región, se evidencia una notable disparidad en el acceso a internet y computadoras. En Chile, Jamaica y Uruguay, casi todos los y las estudiantes tienen acceso a internet, mientras que en Paraguay y Guatemala más de la mitad no tiene acceso a internet. Esto refleja una importante inequidad entre los países con más y menos acceso a esta infraestructura digital de la región. Similarmente, en El Salvador, Costa Rica y Chile casi todos los y las estudiantes tienen acceso a computadoras, mientras que en Paraguay 3 de cada 10 estudiantes no tienen acceso a computadoras, y en República Dominicana y Guatemala 2 de cada 10 estudiantes carecen de acceso (Figura 5 y Figura 6).

Figura 5: acceso a computadoras por país (2022)



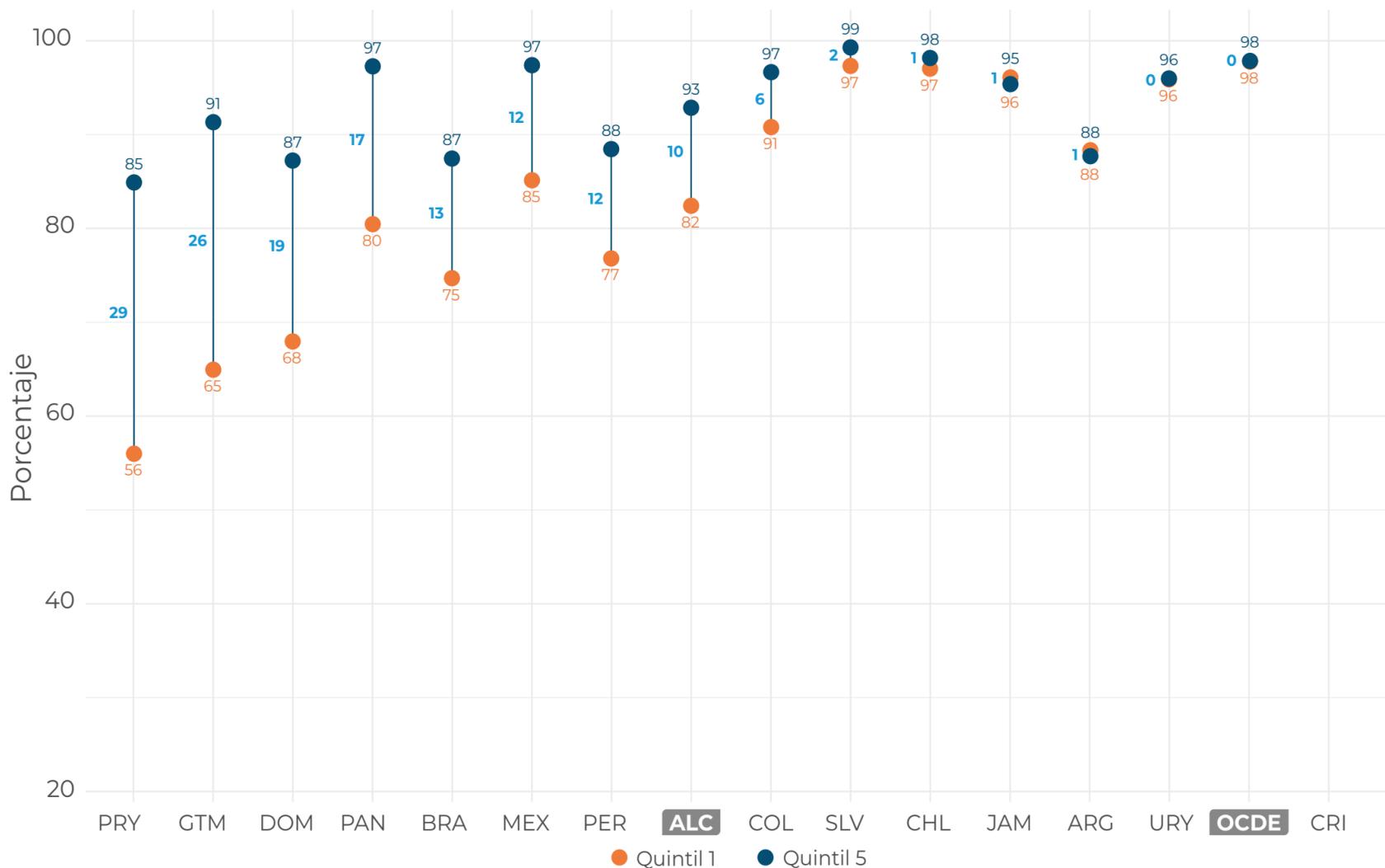
Fuente: elaboración propia en base a CIMA.

Figura 6: acceso a internet por país (2022)



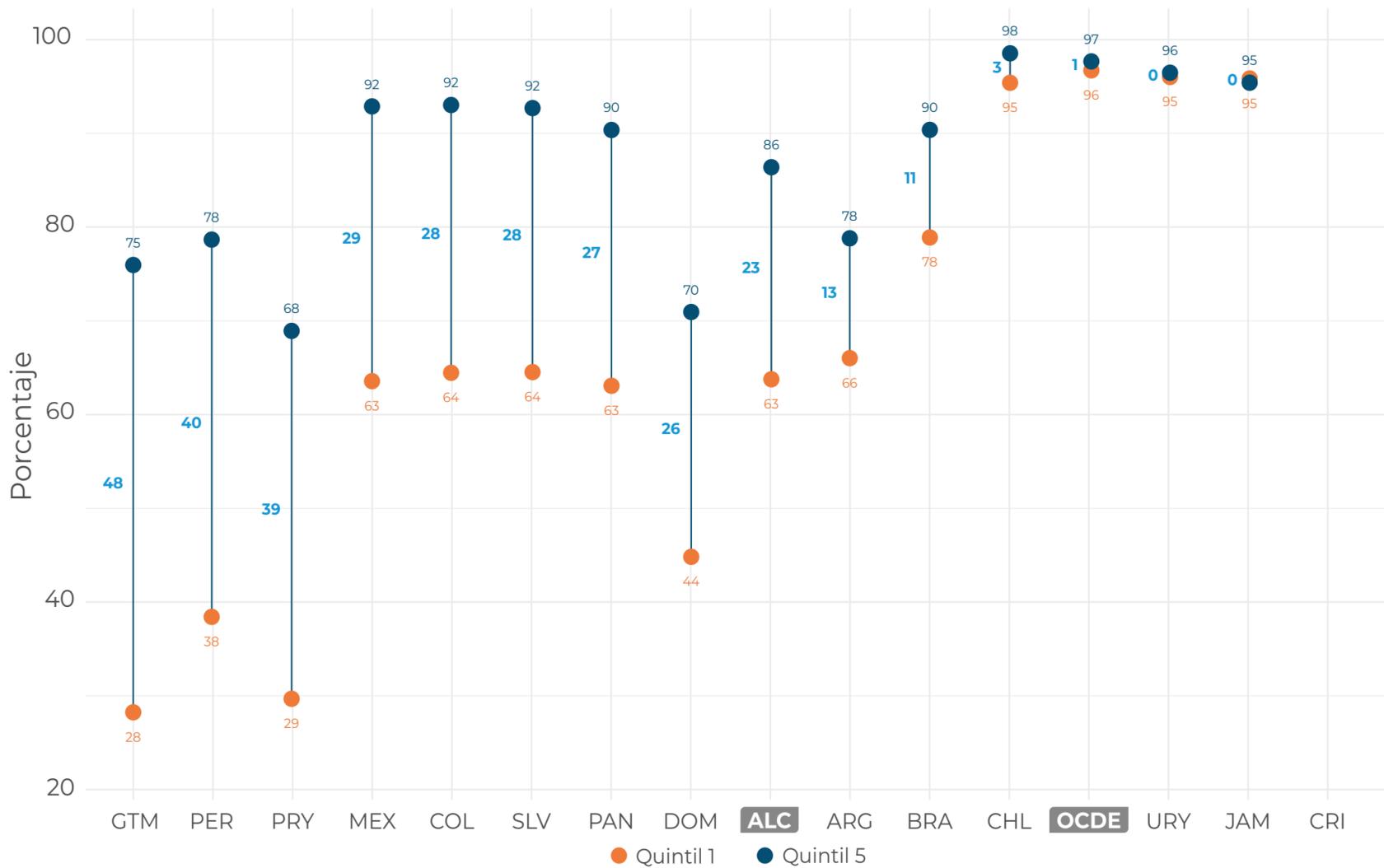
Fuente: elaboración propia en base a CIMA.

Figura 7: acceso a computadora en la institución educativa por nivel socioeconómico (2022)



Fuente: elaboración propia en base a CIMA.

Figura 8: acceso a internet en la institución educativa por nivel socioeconómico (2022)



Fuente: elaboración propia en base a CIMA.

Además de la diferencia de acceso entre países, dentro de los países de América Latina y el Caribe hay una notoria inequidad en el acceso a infraestructura digital por nivel socioeconómico. En general, los países con menor acceso promedio a internet y a computadoras presentan una mayor brecha por nivel socioeconómico.

En Guatemala 2 de cada 10 estudiantes del nivel socioeconómico más alto no tienen acceso a internet, en comparación con 7 de cada 10 del nivel socioeconómico más bajo. En Perú 2 de cada 10 estudiantes del nivel socioeconómico más alto no tienen acceso a internet, frente a 6 de cada 10 del nivel socioeconómico más bajo. Esto contrasta con el acceso completo a internet, indiferente al nivel socioeconómico, que presentan otros países de la región, incluyendo a Chile y Jamaica (Figura 8).

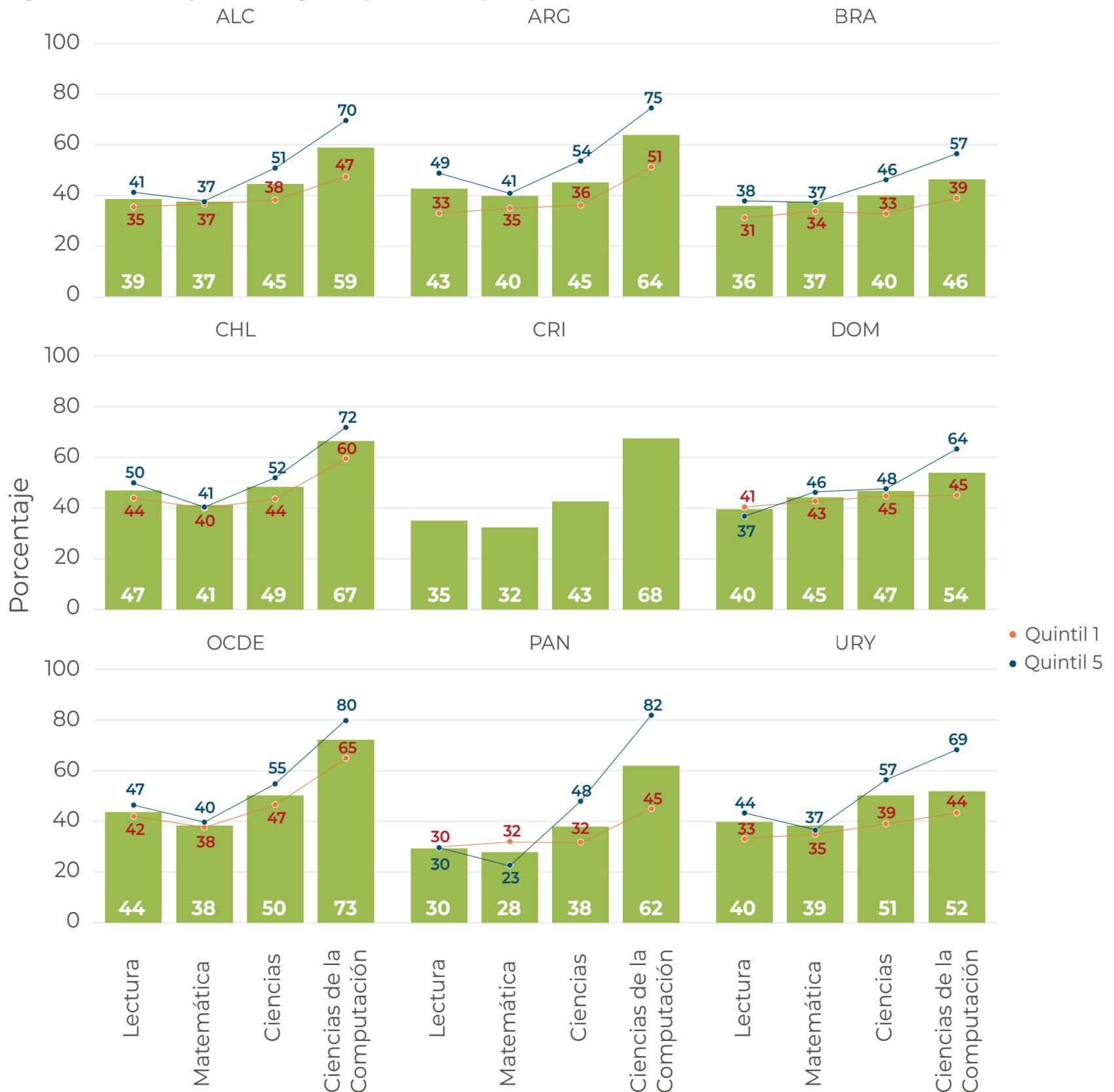
En el acceso a computadoras también se puede observar una brecha por nivel socioeconómico. En Paraguay 1 de cada 10 estudiantes del nivel socioeconómico más alto no tiene acceso a computadoras en la escuela, frente a 4 de cada 10 del nivel más bajo. En Guatemala 1 de cada 10 estudiantes del nivel socioeconómico más alto no tiene acceso a computadoras, mientras que 3 de cada 10 del nivel más bajo no tienen acceso (Figura 7). El deficiente acceso a infraestructura digital, especialmente para estudiantes de contextos socioeconómicos desfavorecidos, ya había sido previamente apuntado (Arias Ortiz et al., 2024).

PISA 2022 también permite analizar el uso de dispositivos digitales por asignatura, según la declaración de las y los estudiantes. La Figura 9

muestra el porcentaje de estudiantes que reporta usar dispositivos digitales en al menos la mitad de las clases de la asignatura analizada. La región presenta una distribución similar al promedio de la OCDE: el mayor uso de dispositivos se observa en clases de computación (59%), luego ciencias (45%), lengua (39%) y por último matemáticas (37%). República Dominicana es el único país que no sigue este patrón, presentando un uso mayor en computación (54%), seguido de ciencias (47%), matemáticas (45%) y lengua (40%) (Figura 9)

Para el promedio de la región, en términos absolutos, la mayor brecha de uso entre el nivel socioeconómico más alto y el más bajo se encuentra en ciencias de computación (23 puntos porcentuales, p. p.), mientras que la brecha más pequeña se observa en matemáticas (0 p. p.). Sin embargo, es importante destacar que no se observan asociaciones claras entre el uso de dispositivos y los resultados de aprendizaje, lo que resalta la importancia de tener una estrategia de implementación eficaz de dichos dispositivos para mejorar los resultados educativos. Muchos países de la región carecen de adecuadas estructuras de gobernanza y directrices a nivel de las instituciones educativas para garantizar el uso eficiente de los recursos digitales para el aprendizaje y la enseñanza (Arias Ortiz et al., 2024). Estas deficiencias subrayan la necesidad de desarrollar e implementar políticas y guías para el personal docente y administrativo de escuelas para potenciar el uso de la tecnología en el ámbito educativo.

Figura 9: uso de dispositivos digitales por materia (2022)



Fuente: elaboración propia en base a CIMA.

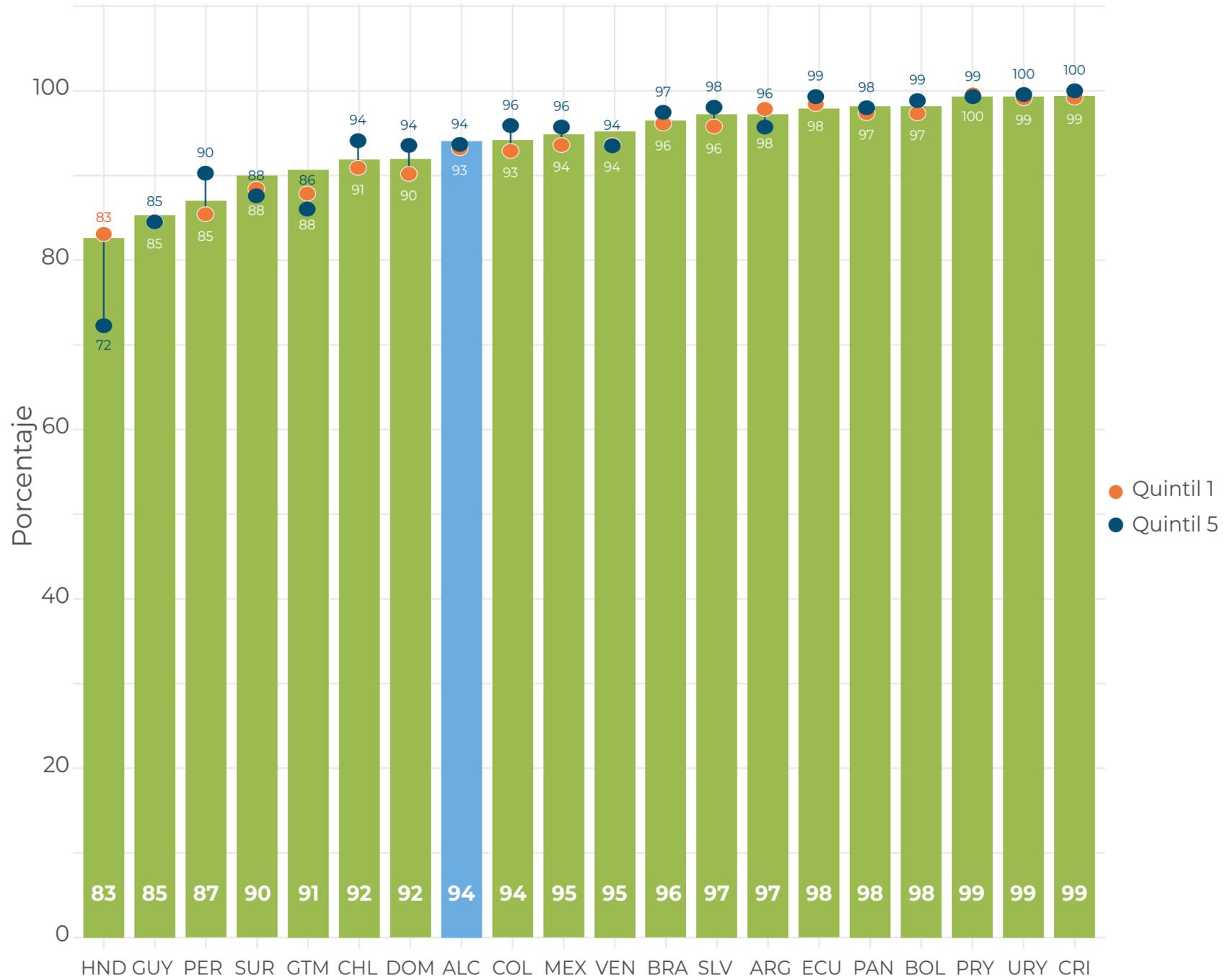
Nota: No hay datos disponibles sobre nivel socioeconómico para Costa Rica en PISA 2022

1.3 Cobertura

El indicador de cobertura mide el porcentaje de la población que está asistiendo efectivamente a un nivel educativo sobre la población que debería estar asistiendo según su edad. En el nivel primario la mayoría de los países de la región ha alcanzado la cobertura universal. Sin embargo, en los niveles secundario y terciario persisten deficiencias, especialmente notables cuando se desagrega la cobertura por nivel socioeconómico.

Aunque la mayoría de los países de América Latina y el Caribe alcanza la cobertura universal en primaria, algunos aún presentan deficiencias. En Honduras 2 de cada 10 personas que deberían estar en primaria no asisten. Esta deficiencia es mayor en el nivel socioeconómico más bajo, donde 3 de cada 10 personas no asisten. En Guyana y Perú 1 de cada 10 personas en edad de asistir a primaria no lo hace. No obstante, es importante señalar que en la mayoría de los países de la región la brecha en la asistencia a primaria entre diferentes niveles socioeconómicos es pequeña o prácticamente inexistente (Figura 10).

Figura 10: brecha en la cobertura neta de educación primaria por nivel socioeconómico (Circa 2023)¹¹



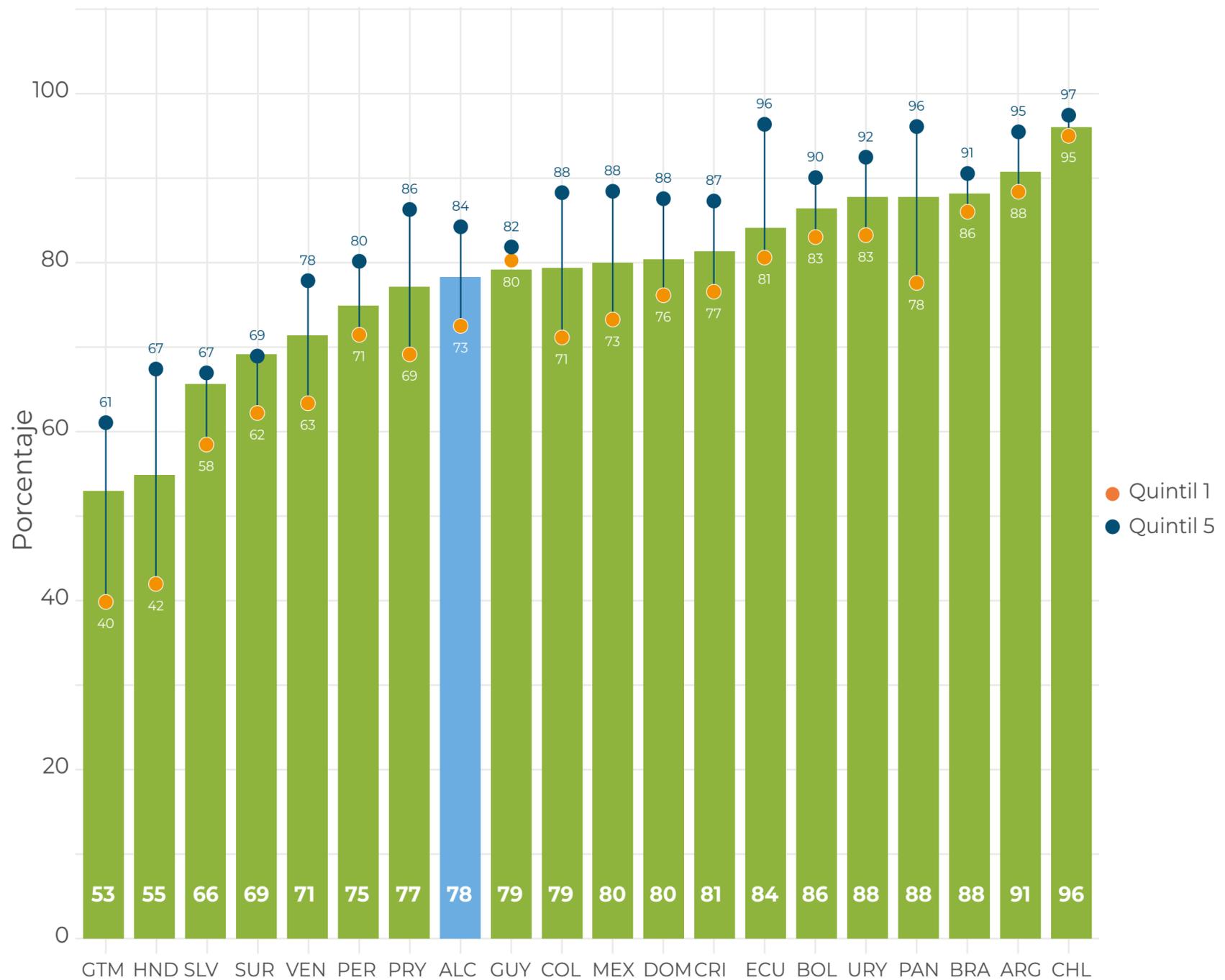
Fuente: elaboración propia en base a CIMA.

En la región 8 de cada 10 personas en edad de asistir al nivel secundario asisten efectivamente al mismo. A diferencia del nivel primario, persisten brechas en la asistencia al nivel secundario entre personas de los niveles socioeconómicos más alto y más bajo (11 p. p.). Los países con las mayores deficiencias en cobertura secundaria son Guatemala, donde 5 de cada 10 personas asisten, y Honduras, con

6 de cada 10. Esta deficiencia se acentúa al analizar la brecha socioeconómica, ya que en ambos países sólo 4 de cada 10 personas en edad de asistir al nivel secundario del nivel socioeconómico más bajo efectivamente asisten. Esto representa una deficiencia del 60 por ciento en comparación con la cobertura universal (Figura 11).

¹¹ Los datos de Bolivia (BOL), República Dominicana (DOM), Guyana (GUY) y Venezuela (VEN) corresponden al año 2021; los datos de Chile (CHL), Argentina (ARG), Brasil (BRA), Panamá (PAN), Uruguay (URY), Costa Rica (CRI), México (MEX), Colombia (COL), Paraguay (PRY), Perú (PER), Surinam (SUR) y Guatemala (GTM) corresponden al año 2022; los datos de Ecuador (ECU), El Salvador (SLV) y Honduras (HND) corresponden al año 2023. El valor para América Latina y el Caribe (ALC) es calculado como promedio simple de los datos de los países.

Figura 11: brecha en la cobertura neta de educación secundaria por nivel socioeconómico (Circa 2023)¹²



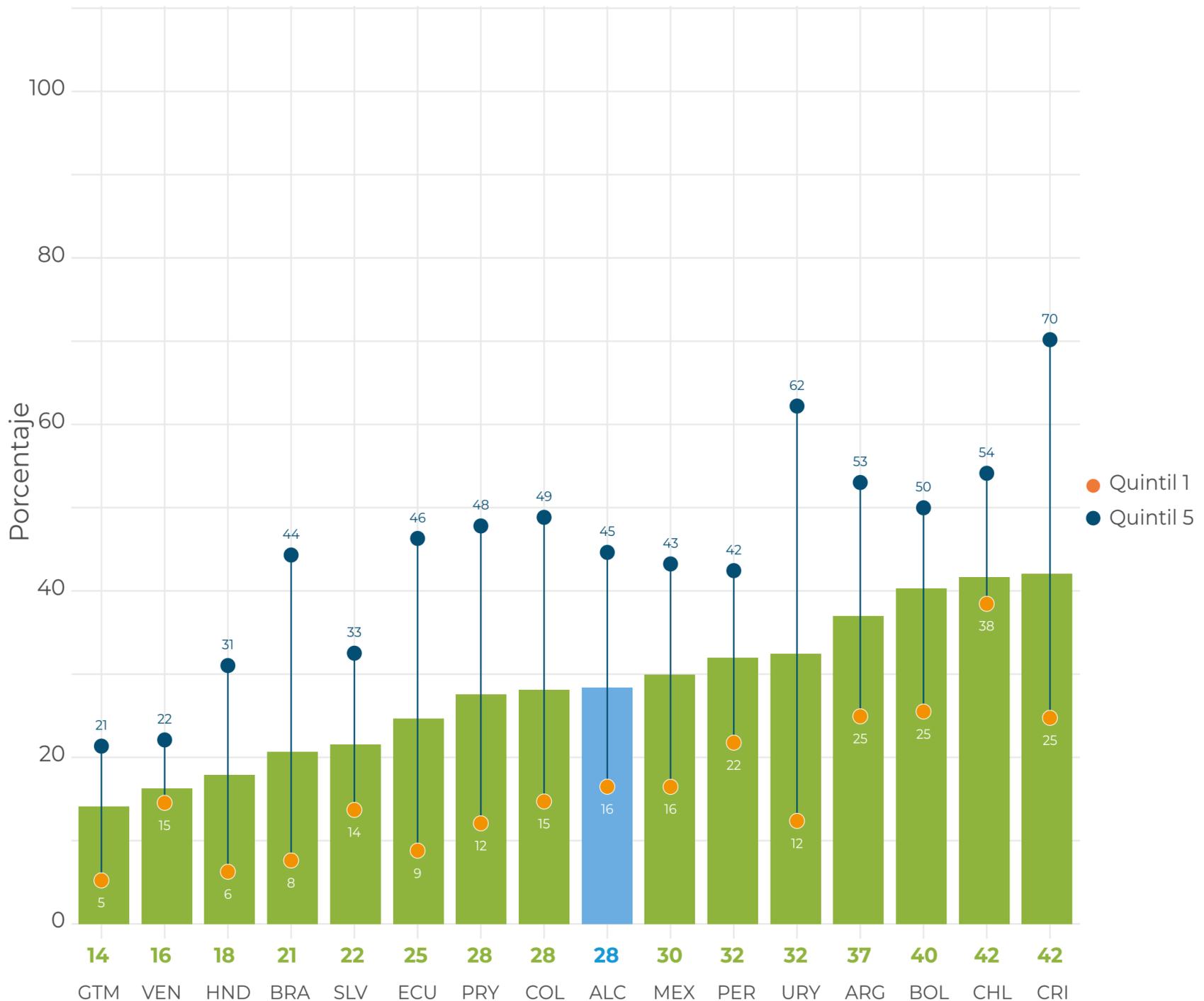
Fuente: elaboración propia en base a CIMA.

La cobertura en educación terciaria es notablemente inferior a la de los niveles primario y secundario. En promedio, sólo 3 de cada 10 personas en edad de asistir a educación terciaria lo hacen, y en todos los países la cobertura es menor al 50 por ciento. Los países con mayor cobertura son Costa Rica, Chile y Bolivia, donde 4 de cada 10 personas asisten. En contraste, los países con menor cobertura en educación terciaria son Guatemala, con 1 de cada 10, y Venezuela y Honduras, con 2 de cada 10 personas en edad de asistir a educación terciaria asistiendo a la misma.

La brecha por nivel socioeconómico de la tasa de asistencia a nivel terciario es mayor a las tasas de los niveles primario y secundario. Los países que presentan las mayores disparidades en la cobertura de educación terciaria entre el nivel socioeconómico más alto y más bajo son Uruguay (50 p. p.), Costa Rica (45 p. p.) y Ecuador (37 p. p.) (Figura 12).

12 Los datos de Bolivia (BOL), República Dominicana (DOM), Guyana (GUY) y Venezuela (VEN) corresponden al año 2021; los datos de Chile (CHL), Argentina (ARG), Brasil (BRA), Panamá (PAN), Uruguay (URY), Costa Rica (CRI), México (MEX), Colombia (COL), Paraguay (PRY), Perú (PER), Surinam (SUR) y Guatemala (GTM) corresponden al año 2022; los datos de Ecuador (ECU), El Salvador (SLV) y Honduras (HND) corresponden al año 2023. El valor para América Latina y el Caribe (ALC) es calculado como promedio simple de los datos de los países.

Figura 12: brecha en la cobertura neta de educación terciaria por nivel socioeconómico (Circa 2023)¹³



Fuente: elaboración propia en base a CIMA.

1.4 Eficiencia

El éxito de un sistema educativo puede ser medido en términos de la eficiencia con la que los y las estudiantes avanzan a través de los distintos niveles educativos hasta su finalización, así como por el nivel de aprendizaje alcanzado. Los indicadores comúnmente utilizados para medir la eficiencia son la tasa de abandono temprano y la tasa de terminación por nivel educativo. Estos indicadores permiten analizar la proporción de estudiantes que ingresan al sistema educativo y completan su formación.

En América Latina y el Caribe 27 por ciento de las y los estudiantes abandonan la educación de forma temprana¹⁴. Los países con las menores tasas de abandono son Chile (7 por ciento), Perú (10 por ciento) y Bolivia (16 por ciento), mientras que las mayores tasas de abandono se registran en Guatemala (57 por ciento), Honduras (53 por ciento) y Uruguay (37 por ciento).

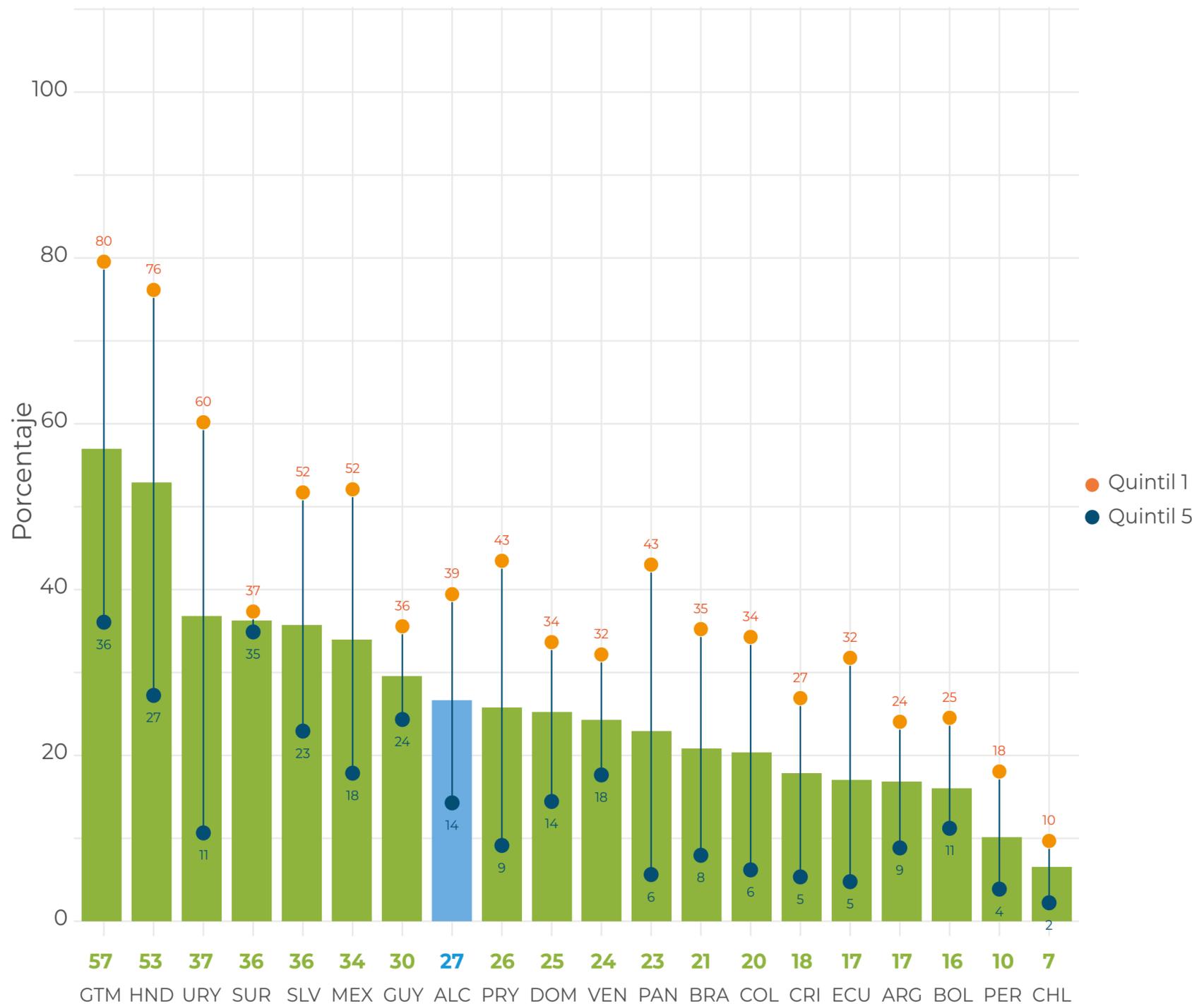
¹³ Los datos de Bolivia (BOL) y Venezuela (VEN) corresponden al año 2021; los datos de Chile (CHL), Argentina (ARG), Brasil (BRA), Uruguay (URY), Costa Rica (CRI), México (MEX), Colombia (COL), Perú (PER), Paraguay (PRY) y Guatemala (GTM) corresponden al año 2022; los datos de Ecuador (ECU), El Salvador (SLV) y Honduras (HND) corresponden al año 2023. El valor para América Latina y el Caribe (ALC) es calculado como promedio simple de los datos de los países.

¹⁴ Abandono temprano es la proporción de jóvenes de entre 18 y 24 años que no han finalizado la educación secundaria y no están asistiendo a ningún nivel de educación.

Se observan significativas desigualdades en las tasas de abandono escolar según el nivel socioeconómico. Efectivamente, en la región 14 por ciento de las y los estudiantes del nivel socioeconómico más alto abandona la educación de forma temprana, frente a 39 por ciento de sus pares del nivel socioeconómico más bajo. Las tasas de abandono más

elevadas para estudiantes del nivel socioeconómico más bajo se registran en Guatemala (80 por ciento), Honduras (76 por ciento) y Uruguay (60 por ciento). Estos países también muestran las mayores brechas en las tasas de abandono entre los niveles socioeconómicos más alto y más bajo (Figura 13).

Figura 13: tasa de abandono temprano por nivel socioeconómico (Circa 2023)¹⁵



Fuente: elaboración propia en base a CIMA.

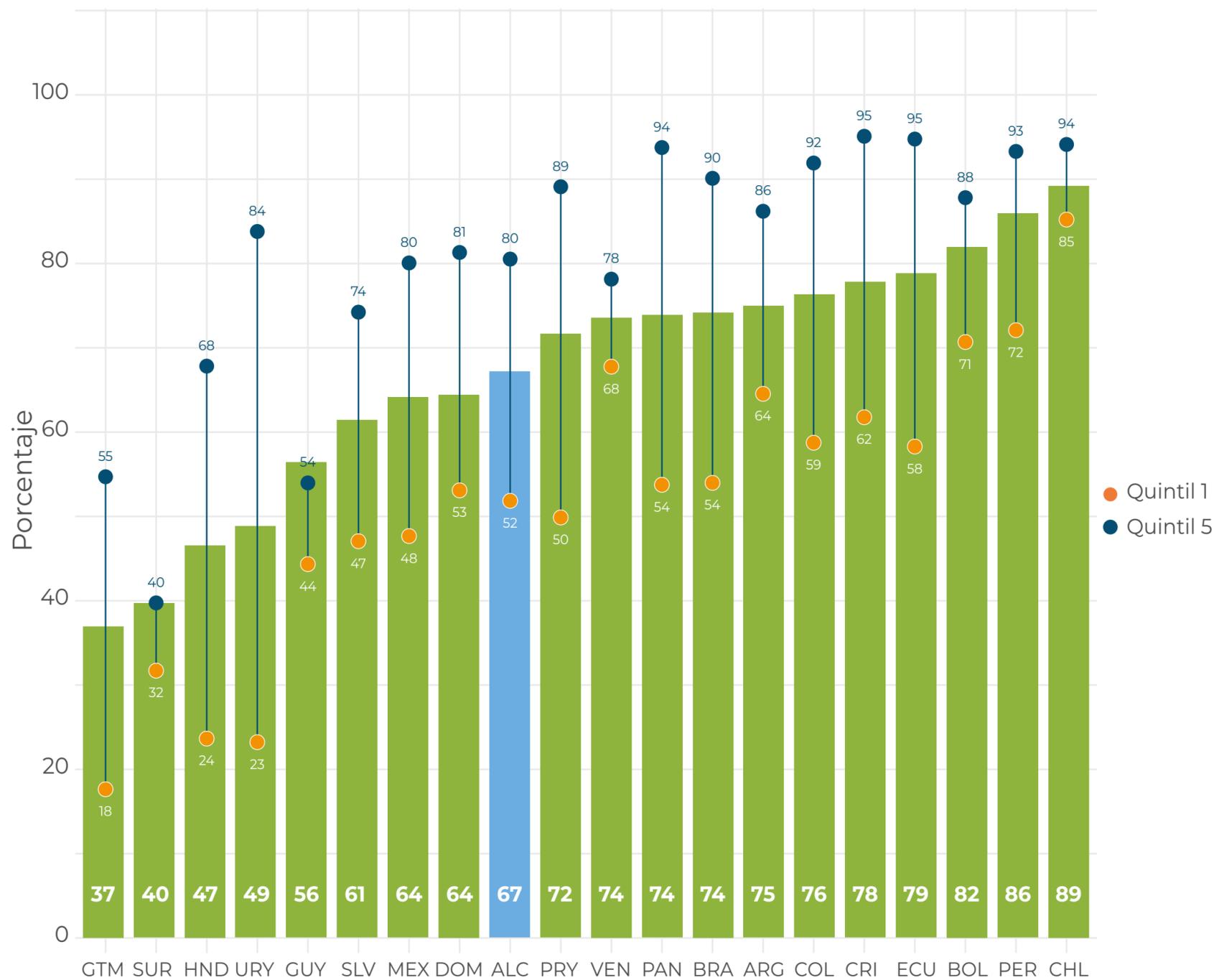
15 Los datos de Bolivia (BOL), República Dominicana (DOM), Guyana (GUY) y Venezuela (VEN) corresponden al año 2021; los datos de Chile (CHL), Argentina (ARG), Brasil (BRA), Panamá (PAN), Uruguay (URY), Costa Rica (CRI), México (MEX), Colombia (COL), Paraguay (PRY), Perú (PER), Surinam (SUR) y Guatemala (GTM) corresponden al año 2022; los datos de Ecuador (ECU), El Salvador (SLV) y Honduras (HND) corresponden al año 2023. El valor para América Latina y el Caribe (ALC) es calculado como promedio simple de los datos de los países.

En la región 67 por ciento de los jóvenes completan la educación secundaria¹⁶. Los países con la mayor eficiencia en este aspecto son Chile (89 por ciento), Perú (86 por ciento) y Bolivia (83 por ciento). En contraste, los países con menor eficiencia son Guatemala (37 por ciento), Surinam (40 por ciento) y Honduras (47 por ciento).

El análisis por nivel socioeconómico resalta las deficiencias de la región en términos de

terminación de la secundaria. Las tasas más bajas de finalización para el nivel socioeconómico más bajo se observan en Guatemala (18 por ciento), Uruguay (23 por ciento) y Honduras (25 por ciento). Esto implica que, en estos países, más de 75 por ciento de una cohorte de jóvenes del nivel socioeconómico más bajo con edades entre 3 y 5 años mayores que la edad teórica para el último grado de educación secundaria no ha completado este nivel (Figura 14).

Figura 14: tasa de terminación de educación secundaria por nivel socioeconómico (Circa 2023)¹⁷



Fuente: elaboración propia en base a CIMA.

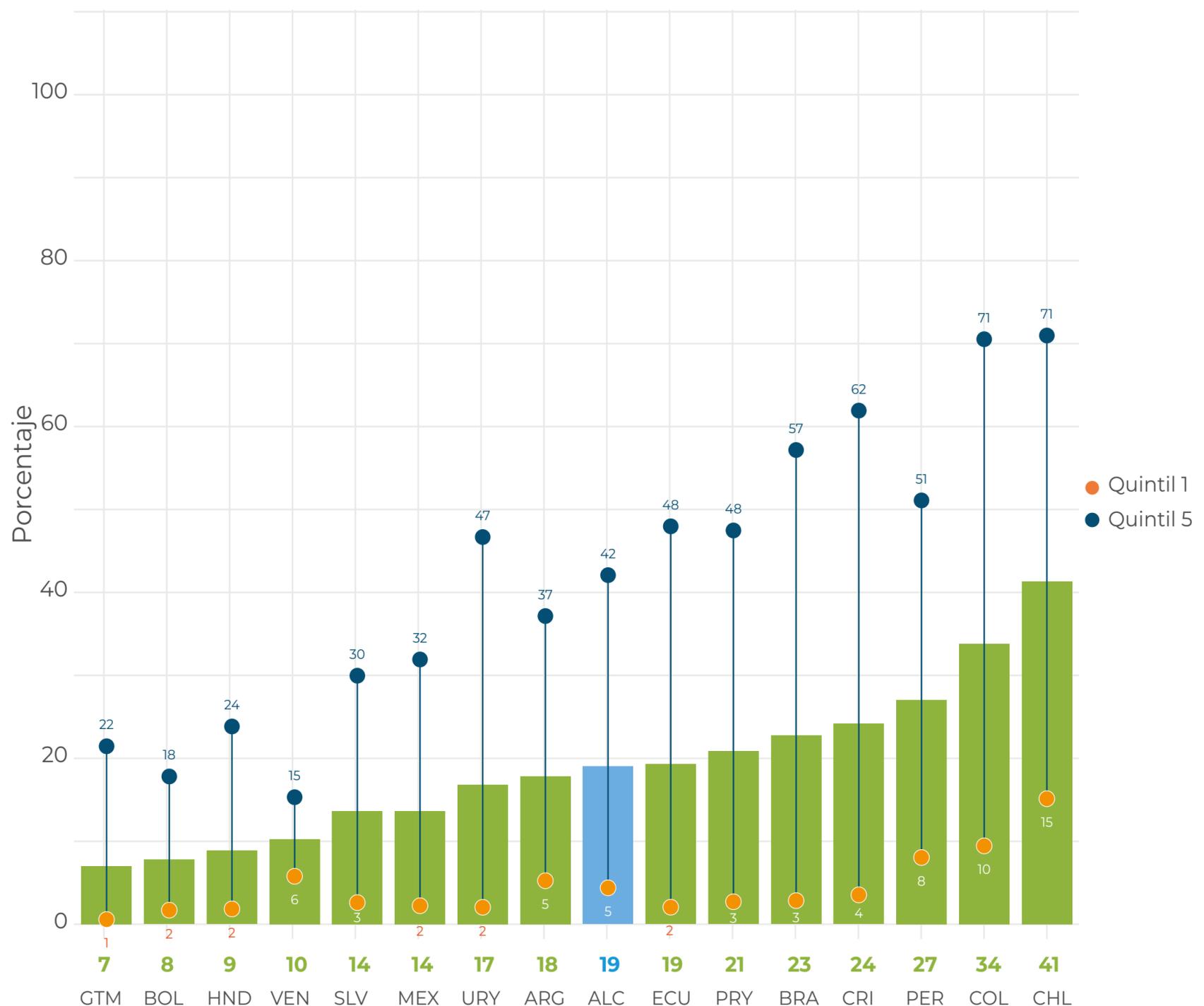
16 La tasa de terminación es el porcentaje de una cohorte de personas con edades entre 3 y 5 años mayores que la edad teórica para el último grado de educación secundaria que ha completado este nivel.

17 Los datos de Bolivia (BOL), República Dominicana (DOM), Guyana (GUY) y Venezuela (VEN) corresponden al año 2021; los datos de Chile (CHL), Argentina (ARG), Brasil (BRA), Panamá (PAN), Uruguay (URY), Costa Rica (CRI), México (MEX), Paraguay (PRY), Perú (PER), Surinam (SUR) y Guatemala (GTM) corresponden al año 2022; los datos de Ecuador (ECU), El Salvador (SLV) y Honduras (HND) corresponden al año 2023. El valor para América Latina y el Caribe (ALC) es calculado como promedio simple de los datos de los países.

En América Latina y el Caribe 19 por ciento de las personas entre 25 a 34 años ha finalizado la educación terciaria¹⁸. Esta baja tasa de finalización en el nivel terciario está vinculada a las bajas tasas de asistencia y terminación en el nivel secundario. Los países con la mayor eficiencia son Chile (41 por ciento), Colombia (34 por ciento) y Perú (27 por ciento). Los países con la menor eficiencia son Guatemala (7 por ciento), Bolivia (8 por ciento) y Honduras (9 por ciento).

Al analizar la brecha en la tasa de terminación de la educación terciaria entre los niveles socioeconómicos más alto y bajo se evidencian significativas inequidades en la región. Mientras que la tasa de finalización en el nivel socioeconómico más alto es de 42 por ciento, en el nivel más bajo es de 5 por ciento, lo que implica una brecha de 37 puntos porcentuales. Además, en la región se observa que a medida que aumenta la tasa promedio de finalización de la educación terciaria también se amplía la brecha entre los niveles socioeconómicos más alto y más bajo (Figura 15).

Figura 15: tasa de terminación de educación terciaria por nivel socioeconómico (Circa 2023)¹⁹



Fuente: elaboración propia en base a CIMA.

18 La tasa de terminación de la educación terciaria es el porcentaje de personas de entre 24 y 35 años que han completado ese nivel educativo

19 Los datos de Bolivia (BOL) y Venezuela (VEN) corresponden al año 2021; los datos de Chile (CHL), Argentina (ARG), Brasil (BRA), Uruguay (URY), Costa Rica (CRI), México (MEX), Colombia (COL), Perú (PER), Paraguay (PRY) y Guatemala (GTM) corresponden al año 2022; los datos de Ecuador (ECU), El Salvador (SLV) y Honduras (HND) corresponden al año 2023. El valor para América Latina y el Caribe (ALC) es calculado como promedio simple de los datos de los países.

I.5 Aprendizaje

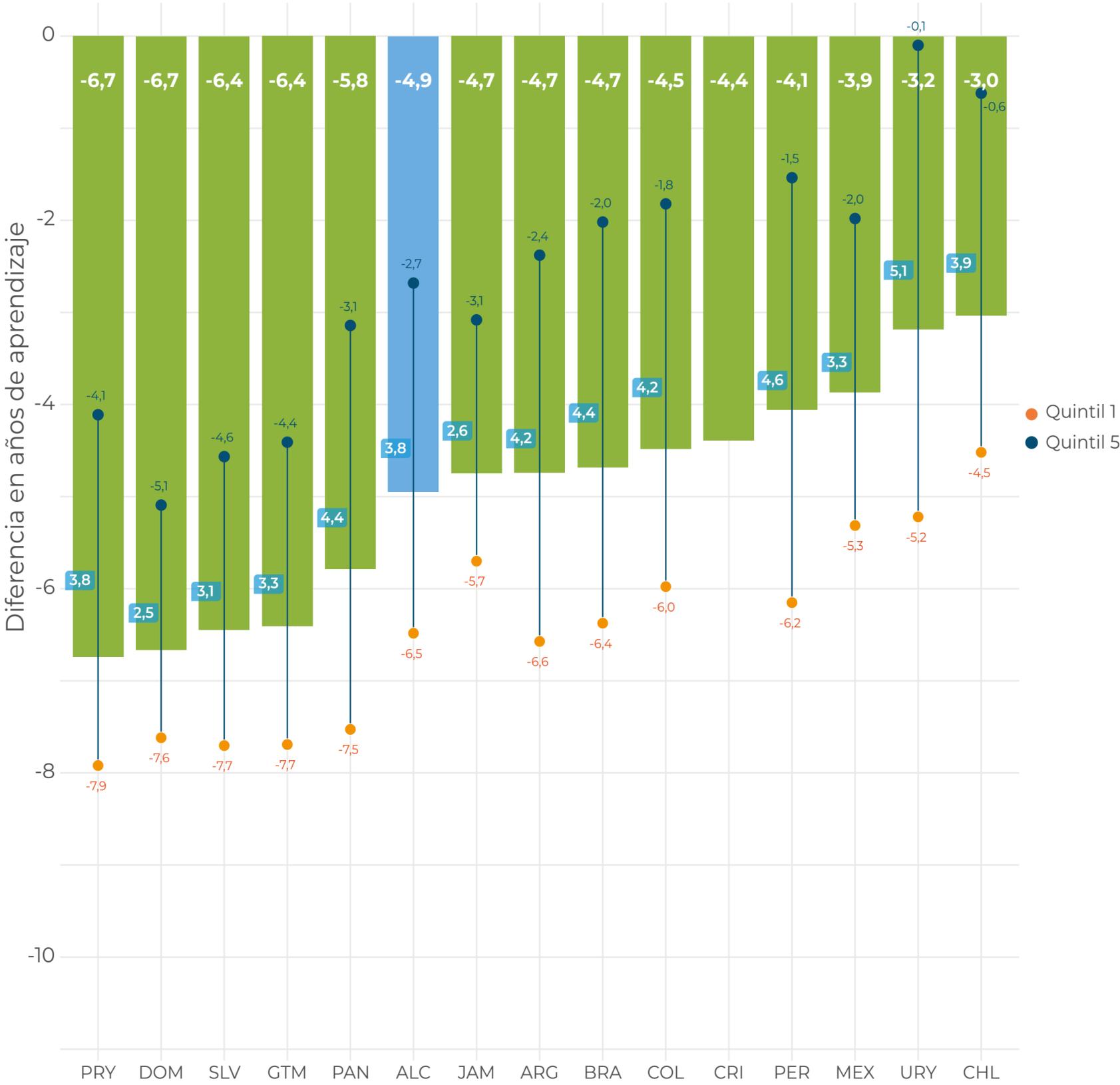
Para analizar el éxito educativo en América Latina y el Caribe en términos de aprendizaje utilizamos los resultados de las pruebas de matemáticas de PISA 2022. Estos resultados han revelado dos grandes desafíos para la región. En primer lugar, un porcentaje considerable de estudiantes no logra desarrollar las competencias fundamentales en matemáticas necesarias para continuar aprendiendo y alcanzar su potencial²⁰. En segundo lugar, persiste una significativa inequidad que desfavorece a las y los estudiantes socioeconómicamente más vulnerables (Arias Ortiz et al., 2023b).

Desde 2018 el acceso a una educación de calidad para las y los estudiantes más pobres de América Latina y el Caribe se ha deteriorado o estancado (Arias Ortiz et al., 2024). Entre 2018 y 2022 la proporción de estudiantes con bajo desempeño en matemáticas entre los más pobres aumentó entre 3 y 8 puntos porcentuales en Brasil, Colombia, México y Panamá. En los demás países que participaron en ambas rondas esta proporción se mantuvo sin cambios durante el período. En cuanto a la tendencia de largo plazo, entre 2012 y 2022 la proporción de estudiantes con bajo desempeño en matemáticas entre los más pobres aumentó o se mantuvo sin cambios en casi todos los países de la región. Específicamente, la prevalencia del bajo desempeño en matemáticas entre las y los estudiantes más pobres de Argentina, Brasil y México aumentó entre 3 y 10 puntos porcentuales (Arias Ortiz et al., 2024).

Un incremento de 20 puntos en una prueba PISA refleja aproximadamente el avance anual promedio de las y los estudiantes que participan en la evaluación (OCDE, 2023). En otras palabras, una puntuación 20 puntos por debajo del promedio de la OCDE equivale a un año de retraso en comparación al promedio de la OCDE (Arias Ortiz et al., 2023b). América Latina y el Caribe presenta, en promedio, un rezago de cinco años en comparación con el promedio de la OCDE, y todos los países exhiben un rezago de al menos tres años. Los países con el mayor rezago respecto al promedio de la OCDE son Paraguay (7 años), República Dominicana (7 años) y El Salvador (6 años). También existen brechas significativas dentro de los países de la región entre estudiantes del nivel socioeconómico más alto y estudiantes del nivel socioeconómico más bajo. En más de la mitad de los países las y los estudiantes del nivel socioeconómico más bajo presentan un rezago de por lo menos seis años respecto al promedio de la OCDE; el mayor rezago se verifica en Paraguay (8 años), El Salvador (8 años) y Guatemala (8 años). Las mayores brechas entre el nivel socioeconómico más alto y el más bajo se encuentran en Uruguay (5,1 años), Perú (4,6 años), Brasil (4,4 años) y Panamá (4,4 años) (Figura 16).

20 En PISA 2022 el dominio principal bajo estudio fue matemáticas, aunque también se evaluaron lectura y ciencias.

Figura 16: rezago en años de escolaridad en comparación al promedio OCDE (2022)



Fuente: elaboración propia en base a CIMA.

PARTE II.**Evolución y perspectivas de la medición de aprendizajes**

Diferentes aspectos de los sistemas escolares –incluyendo la competencia de las y los maestros, las características de las escuelas y las políticas educativas– pueden optimizarse para mejorar las oportunidades de aprendizaje de grupos de estudiantes en diferentes circunstancias. El objetivo de la investigación y la política pública en educación es entender la combinación de factores que, en un contexto específico, hace que un sistema educativo sea eficiente, efectivo y equitativo (Clarke y Lune-Bazaldua, 2021). La implementación de evaluaciones a gran escala, tanto a nivel nacional como internacional, ha abierto numerosas oportunidades de investigación que ayudan a comprender mejor cómo ha contribuido esta combinación de factores, traducida en programas y políticas públicas, a la mejora de los sistemas educativos en diferentes países (Teig y Steinmann, 2023).

¿Por qué hablar de evaluaciones de aprendizaje? En primer lugar, las evaluaciones nacionales a gran escala permiten medir el desempeño de los y las estudiantes en relación con los objetivos de aprendizaje establecidos en el currículo nacional o los lineamientos curriculares. Estas evaluaciones son una herramienta esencial para monitorear la calidad de los sistemas educativos, enfocándose en los resultados de aprendizaje, en lugar de otros indicadores que miden insumos, acceso o finalización. A medida que la discusión de la política educativa se centra cada vez más en la calidad y la equidad del aprendizaje aumenta el uso de estas evaluaciones para monitorear los resultados. Esto refleja un cambio significativo desde el uso inicial de las evaluaciones, que buscaba contrastar el conocimiento teórico sobre procesos educativos con evidencia empírica, hacia un enfoque más centrado en

evaluar la calidad de los sistemas educativos (Wagemaker, 2014). Además, las evaluaciones a gran escala pueden proporcionar información sobre la relación entre el contexto, los recursos y el desempeño, lo que permite monitorear la equidad en el aprendizaje y, en el mejor de los casos, promover políticas que fomenten una distribución más equitativa de las oportunidades educativas (Arregui, 2008). En este sentido, las evaluaciones a gran escala son frecuentemente utilizadas por los responsables de políticas para diagnosticar problemas y dirigir intervenciones a grupos desfavorecidos.

Junto con su rol en el monitoreo de la calidad y la equidad educativa, las evaluaciones de aprendizaje desempeñan un papel crucial en la rendición de cuentas y en el apalancamiento de políticas. Estas evaluaciones permiten a los y las responsables de políticas garantizar la transparencia al informar los resultados a las partes interesadas y fortalecen la confianza en los sistemas educativos. Por su parte, el apalancamiento se refiere a la capacidad de utilizar los resultados de las evaluaciones para impulsar prioridades de política, creando conciencia entre los diversos actores sobre aspectos críticos de la calidad educativa o la magnitud de un problema de reforma que haya sido identificado. Ambos objetivos, rendición de cuentas y apalancamiento, subrayan la importancia de las evaluaciones como instrumentos estratégicos en la formulación y gestión de las políticas educativas.

Las evaluaciones de aprendizaje deben consolidarse como herramientas eficientes para recopilar y proporcionar evidencia que informe la toma de decisiones en política educativa. En las últimas décadas, varios países de América Latina y el Caribe han fortalecido sus sistemas de evaluación de aprendizaje, generando beneficios concretos en términos de política pública. En este contexto, es fundamental comprender la relación entre las evaluaciones a gran escala y el ciclo de políticas.

Tobin, Nugroho, y Lietz (2016) ofrece una revisión exhaustiva que muestra cómo, en países de ingresos medios y bajos, las evaluaciones se utilizan con mayor frecuencia en las etapas de definición de la agenda, implementación de políticas y monitoreo y evaluación, y con menor frecuencia en la formulación de políticas. Aunque es indiscutible que las evaluaciones nacionales son esenciales para el monitoreo y la evaluación de políticas, es crucial que tengan un impacto significativo en la formulación de políticas educativas. Es posible identificar cuatro áreas clave en las que las evaluaciones pueden incidir en las políticas educativas: la reforma curricular y los estándares de rendimiento, las políticas de evaluación, la asignación de recursos y las políticas de enseñanza y aprendizaje (Tobin, Nugroho y Lietz, 2016).

Diversos ejemplos de la región ilustran cómo las evaluaciones nacionales de aprendizaje han desempeñado un papel central en la formulación de políticas en estas áreas. En el ámbito de la reforma curricular y los estándares de rendimiento, Colombia (2002), Guatemala (2005) y Argentina (2004) implementaron procesos de ingeniería inversa en los que las evaluaciones impulsaron el desarrollo de estándares de aprendizaje más claros y medibles al estructurar pruebas normativas²¹ (Arregui, 2008). Chile es un ejemplo destacado en la asignación de recursos. La Ley SEP, en la promulgada en 2008, utiliza los resultados de las evaluaciones para clasificar a las escue-

las según su desempeño, asignar recursos y definir el nivel de apoyo y autonomía según su clasificación. Esta reforma, orientada a la equidad, se basó en la publicación de las persistentes brechas de aprendizaje, lo cual fue clave para impulsar la redistribución de recursos hacia las escuelas más pobres (Galas, Gutiérrez y Hamilton, 2022).

En lo que respecta a las políticas de enseñanza y aprendizaje, si se diseminan efectivamente en las escuelas, los resultados de las evaluaciones nacionales pueden orientar cambios en las prácticas de aula y mejorar la formación docente. En Argentina, estudios como el de Hoyos, Ganimian y Holland (2019) demuestran que la retroalimentación diagnóstica proporcionada a los maestros sobre los resultados de las evaluaciones está asociada a un mejor rendimiento estudiantil, pues las y los maestros dedican más tiempo a la instrucción y emplean más actividades de aprendizaje en el aula. De manera similar, cuando las y los directores tienen acceso a esta retroalimentación es más probable que utilicen los resultados para tomar decisiones de gestión, tales como establecer metas de aprendizaje a nivel escolar, actualizar el currículo y tomar decisiones sobre el personal (Clarke y Lune-Bazaldua, 2021). Un estudio reciente en Haití revela que compartir información sobre los resultados con directores en zonas rurales tuvo un impacto positivo en la gestión escolar y en los aprendizajes de las y los estudiantes (Borger et al., 2024).

Ahora bien, para que las evaluaciones de aprendizaje influyan efectivamente en el ciclo de políticas educativas deben cumplirse ciertas condiciones. Según Tobin, Nugroho y Lietz (2016), existen varias barreras clave que pueden obstaculizar este proceso: (a) problemas con la calidad y el análisis de los programas de evaluación; (b) limitaciones financieras para llevar a cabo evaluaciones de alta calidad o para mantener programas continuos; (c) agencias de evaluación débiles que sean vul-

21 En evaluaciones con base normativa, los resultados se obtienen mediante la comparación de los resultados de los estudiantes con un grupo de referencia, lo que impide realizar comparaciones a menos que se utilicen pruebas equiparadas. Por otro lado, la evaluación con base criterial se basa en la comparación de los resultados de los estudiantes con criterios predefinidos, generalmente derivados del currículo, permitiendo determinar el nivel de rendimiento en función de estándares externos previamente establecidos (Murillo, 2007).

nerables a cambios de régimen y que tengan relaciones frágiles con el gobierno; y (d) una capacidad técnica insuficiente del personal para interpretar y comunicar los resultados de manera efectiva. En la región estas barreras son más comunes de lo que sería deseable.

Como en otros aspectos de la política educativa, el cierre de escuelas debido al COVID-19 tuvo un efecto disruptivo. En algunos países de la región las evaluaciones guiaron la priorización de acciones durante la educación remota de emergencia, especialmente las de carácter censal. En México los resultados de las pruebas PLANEA se utilizaron para definir el currículo y los materiales educativos, analizando los contenidos que generalmente resultan más difíciles para las y los estudiantes y estructurando guías pedagógicas para las y los docentes durante la pandemia. En otros países jugaron un papel central en la identificación de la magnitud de las pérdidas de aprendizaje y en la generación de acciones para abordarlas. En Colombia la continuidad de las pruebas Saber 11 permitió medir el impacto del COVID-19 en los aprendizajes de las y los estudiantes del último grado de educación secundaria y conocer las características de las escuelas con mayores tasas de abandono (Dueñas et al., 2022). En otros países las evaluaciones de aprendizaje pasaron a segundo plano, no siendo priorizadas como elemento central ante la situación de emergencia.

Para entender mejor cual ha sido el papel de las pruebas de aprendizaje en la definición de las políticas educativas de los países de la región, en las siguientes secciones analizamos la evolución de los sistemas de evaluación nacionales de los países en la región, tanto a nivel de su participación en pruebas internacionales y regionales como de las pruebas nacionales que han implementado.

II.1 Evaluaciones internacionales y regionales

Aunque las evaluaciones a gran escala en educación se utilizan en estudios comparativos desde hace 60 años, la generalización y discusión global de sus resultados se han desarrollado principalmente en las últimas dos décadas (Bruns et al., 2019). Desde la segunda mitad de la década de 1990, América Latina ha experimentado un notable aumento en el número de países que implementan pruebas de aprendizaje a gran escala a nivel nacional. Para comprender la evolución y consolidación de los sistemas nacionales de aprendizaje en América Latina y el Caribe es ineludible tener en cuenta las evaluaciones internacionales y regionales de aprendizaje.

Concebidas inicialmente para explorar la variación transnacional en las instituciones y procesos educativos y su relación con los resultados de aprendizaje (Keeves, 1995), las evaluaciones regionales e internacionales permiten comparar el rendimiento entre países y monitorear la calidad del aprendizaje a nivel global (Vegas y Petrow, 2008). La iniciativa Educación para Todos (EFA, por sus siglas en inglés) fue lanzada en 1990 durante la Conferencia Mundial sobre Educación para Todos en Tailandia. Liderada por la UNESCO, EFA ha sido un catalizador significativo para el avance de las evaluaciones internacionales y regionales, colocando a la calidad educativa y a la necesidad de su monitoreo constante en el centro del debate educativo global. La EFA promovió el desarrollo e implementación de evaluaciones comparativas internacionales a nivel global y la consolidación de evaluaciones regionales como ERCE (Estudio Regional Comparativo y Explicativo) en América Latina. La evidencia da cuenta de la incidencia de este tipo de evaluaciones en la política educativa nacional, con atributos que complementan las funciones de los sistemas nacionales de evaluación. En países de ingresos medios y bajos, la participación en este tipo de evaluaciones ha tenido una incidencia preponderante en la creación o reforma de las agencias y programas de evaluación nacionales (Tobin,

Nugroho y Lietz, 2016). Además, la participación en este tipo de evaluaciones ha sido un mecanismo para desarrollar y fortalecer capacidades técnicas y metodológicas a nivel local ya que en el proceso de preparación y aplicación se abren oportunidades para el intercambio técnico y para adquirir conocimiento de metodologías de medición cada vez más complejas (Ferrer y Fiszbein, 2015). La participación en este tipo de evaluaciones ha enriquecido y desarrollado las capacidades de los equipos nacionales de expertos (Cox y Meckes, 2016).

Finalmente, las evaluaciones internacionales y regionales, y especialmente PISA, se destacan por su influencia en la definición de agendas educativas en América Latina, sirviendo como un poderoso mecanismo para apalancar prioridades de política educativa. Según la literatura, estas pruebas tienen impacto en la configuración de la agenda educativa cuando se utilizan comparaciones internacionales para desencadenar debates sobre la calidad educativa, particularmente cuando los resultados son más bajos de lo esperado (Tobin, Nugroho y Lietz, 2016).

Evaluaciones regionales

En 1994 se creó el Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (LLECE) de la Oficina Regional de Educación para América Latina y el Caribe (OREALC) de la UNESCO, lo que constituye un hito para la evaluación educativa en los países de habla hispana y portuguesa de la región. LLECE surgió cuando muchos países de la región estaban adoptando reformas educativas sin información suficiente y relevante para su diseño y carecían de una masa crítica de recursos para medir la calidad de la educación (Vanni y Valenzuela, 2020). Hacia 1994, sólo seis países

de América Latina –Argentina, Brasil, Chile, Colombia, Nicaragua y República Dominicana– contaban con mediciones de aprendizaje a gran escala, y sólo tres de ellos (Argentina, Brasil y Chile) contaban con experiencias de evaluación de aprendizajes a nivel de la educación primaria.

Los programas de evaluación regionales se establecieron, en sintonía con la iniciativa EFA, para realizar evaluaciones comparativas entre países con similares contextos geográficos, culturales, lingüísticos y económicos; eso las hace especialmente valiosas para los países de ingresos medios y bajos (Tobin, Nugroho y Lietz, 2016).

La primera versión del estudio regional se llamó PERCE (1997), incorporó a doce países de América Latina y el Caribe y evaluó lectura y matemáticas en tercer y cuarto grado²². Las pruebas de aprendizaje de LLECE se basan en una revisión de los programas curriculares de los países participantes en cada ciclo; se busca identificar el contenido común entre ellos y sobre ese contenido se diseñan las pruebas. Las subsiguientes rondas, SERCE 2006 y TERCE 2013, consolidaron la participación de quince países²³, evaluaron tercer y sexto grado de primaria e incorporaron las áreas de ciencias y escritura en sexto grado. En el cuarto Estudio Regional Comparativo y Explicativo (ERCE 2019) participaron también quince países, pero sumándose por primera vez El Salvador y sin la participación de Chile²⁴. En ERCE 2019 se evaluó el desempeño de las y los estudiantes de tercer y sexto grado en matemáticas, lectura y escritura; además, se evaluó el desempeño en ciencias naturales únicamente en sexto grado. También se fortaleció el conjunto de cuestionarios dirigidos a estudiantes, docentes, familias y directores de escuela. ERCE 2019 también incorporó un nuevo módulo

22 Argentina, Bolivia, Brasil, Chile, Colombia, Costa Rica, Cuba, Honduras, México, Paraguay, Perú, República Dominicana y Venezuela. Aunque en total participaron trece países, nuestro análisis se limita a los veintiséis países miembros del BID, por lo que no se contabilizan los datos de Cuba. Este criterio se aplica también para las siguientes rondas.

23 En SERCE 2006 participaron Argentina, Brasil, Chile, Colombia, Costa Rica, Cuba, Ecuador, El Salvador, Guatemala, México, Nicaragua, Panamá, Paraguay, Perú, República Dominicana y Uruguay. Los países participantes en TERCE 2013 fueron Argentina, Brasil, Chile, Colombia, Costa Rica, Ecuador, Guatemala, Honduras, México, Nicaragua, Panamá, Paraguay, Perú, República Dominicana y Uruguay.

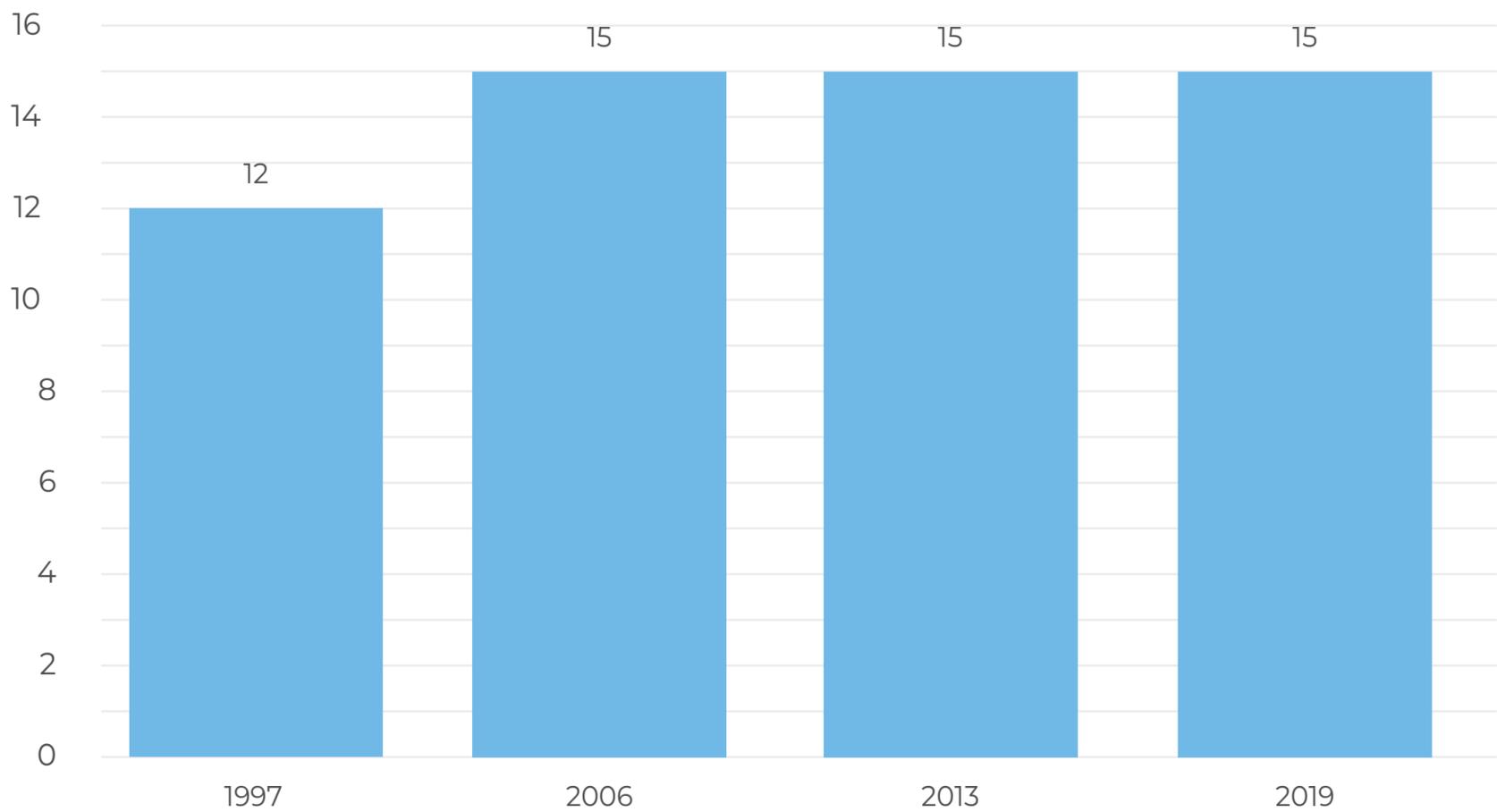
24 Argentina, Brasil, Colombia, Costa Rica, Cuba, Ecuador, El Salvador, Guatemala, Honduras, México, Nicaragua, Panamá, Paraguay, Perú, República Dominicana y Uruguay.

para medir las habilidades socioemocionales de las y los estudiantes de sexto grado, con un enfoque específico en la conciencia y valoración del otro, la autorregulación y el auto-manejo (Vanini y Valenzuela, 2020).

Con un nivel de participación alto y estable desde mediados de la primera década del

siglo, las evaluaciones de LLECE continúan siendo la evaluación regional más importante para los países de América Latina de habla castellana y portuguesa (Brasil), pues no incluye a los países anglófonos ni francófonos de la región (Vanni y Valenzuela, 2020).

Figura 17: cantidad de países participantes en ERCE



Fuente: elaboración propia sobre la base de datos de LLECE.

El papel del CXC en las evaluaciones educativas del Caribe anglófono

El Consejo de Exámenes del Caribe (CXC, por sus siglas en inglés) se ha consolidado como una referencia crucial para las evaluaciones de aprendizaje en el Caribe de habla inglesa²⁵. El CXC ofrece un amplio abanico de evaluaciones a gran escala en distintos niveles educativos, brinda apoyo para el desarrollo de programas de estudio y ofrece servicios a instituciones educativas en el diseño, desarrollo y administración de evaluaciones (Caribbean Examinations Council, s.f.).

Establecido en 1972 por un acuerdo entre gobiernos de quince territorios del Caribe de habla inglesa, el CXC surgió con la misión de desarrollar currículos y sistemas de exámenes culturalmente relevantes en un contexto postcolonial que enfatizaba la importancia de la independencia cultural, sin perder de vista la necesidad de ofrecer certificaciones competitivas a nivel internacional (Gordon, 2019). Fue la primera iniciativa transnacional en la región con competencias sobre el desarrollo e implementación de evaluaciones a gran escala. Desde entonces, el CXC ha evolucionado para incluir a dieciséis países participantes del Caribe y cuatro territorios externos (Caribbean Examinations Council, 2023)²⁶.

Los primeros esfuerzos del CXC se concentraron en el desarrollo del Certificado de Examen Secundario del Caribe (CSEC, por sus siglas en inglés). El CSEC es una evaluación creada en 1979 que proporcionó una nueva hoja de ruta regional y culturalmente relevante, reemplazando al sistema de GCE O-Level, administrado por las autoridades educativas británicas, como el estándar de evaluación de las y los estudiantes al final de la educación secundaria (Gordon, 2019). El CSEC abarca competencias generales y técnicas, proporcionando a los y las estudiantes la acreditación para sus estudios terciarios y su entrada al mundo laboral (Thailinger et al., 2023).

Casi medio siglo después, el CSEC ha consolidado su participación y ampliado las competencias y materias evaluadas. Actualmente evalúa 33 materias, de las cuales 28 corresponden a competencia general y cinco a competencia técnica. El esquema de calificación, dividido en seis niveles, informa sobre el rendimiento del estudiante. Aunque la graduación del nivel secundario no depende de rendir la evaluación, el desempeño en el CSEC es un factor determinante para la progresión hacia la educación terciaria y para las oportunidades de empleo futuro: para ser admitido en la educación terciaria se requiere haber aprobado al menos cinco materias, incluyendo inglés y matemáticas; y los puestos en el sector público también exigen la aprobación de al menos cinco materias del CSEC (Caribbean Examinations Council, s.f.).

El CXC ha desarrollado una variedad de evaluaciones y certificaciones para atender las necesidades de la región a nivel de educación básica, aunque con menos alcance y participación. Entre ellas se destacan el Certificado de Competencia de Nivel Secun-

25 Los primeros antecedentes del CXC se remontan a 1946, mucho antes de que las evaluaciones internacionales y regionales se convirtieran en un tema destacado en la agenda internacional. Casi dos décadas después, durante la Conferencia de Directores de Escuelas Secundarias del Caribe, surgió un sólido respaldo para establecer un consejo de exámenes para la región (Gordon, 2019).

26 Seis de los países participantes del CXC son miembros del BID: Barbados, Belice, Guyana, Jamaica, Trinidad y Tobago y Surinam.

dario del Caribe (CCSLC, por sus siglas en inglés) en 2007, la Calificación Vocacional del Caribe (CVQ, por sus siglas en inglés) en 2008 y la Evaluación de Salida de Primaria del Caribe (CPEA, por sus siglas en inglés) en 2012²⁷.

El impacto de las evaluaciones lideradas por el CXC en el Caribe ha sido significativo, fortaleciendo las capacidades técnicas de los países participantes. Esto ha sido posible gracias a una estrecha colaboración con expertos y organizaciones de otras regiones del mundo, lo que ha permitido mantener estándares de calidad en las evaluaciones. A diferencia del ERCE, el CSEC es una evaluación de alta implicancia (*“high-stakes”*)²⁸ ya que tiene como objetivo certificar conocimientos y es determinante para ingresar a estudios

terciarios y al sector público. A pesar de estas características, el CSEC sigue siendo la principal fuente de información transnacional de los países del Caribe de habla inglesa para comprender la calidad del aprendizaje de las y los estudiantes al finalizar la educación secundaria.

La experiencia del CXC puede servir como un sólido cimiento para futuras colaboraciones y mejoras en la evaluación a gran escala en América Latina y el Caribe. En este sentido, resulta crucial construir puentes entre las diversas iniciativas regionales, en particular con LLECE, para fortalecer y enriquecer las prácticas de evaluación en toda la región.

27 El CPEA es una prueba que toman las y los estudiantes de último año de primaria de algunos países del Caribe antes de avanzar hacia la educación secundaria; ningún país miembro del BID participa de la misma

28 Tomando la definición de Herrero et al. (2022), las pruebas de altas implicancias incluyen aquellas en las que las consecuencias asociadas a sus resultados son vinculantes para distintos actores del sistema (estudiantes, escuelas o maestros). Estas pueden ser: pruebas con fines de certificación de un curso o nivel educativo (primario o secundario) o de selección o admisión para los niveles superiores; o evaluaciones con fines de monitoreo, donde los resultados se utilizan para la rendición de cuentas por parte de las instituciones educativas o existen incentivos asociados a los resultados.

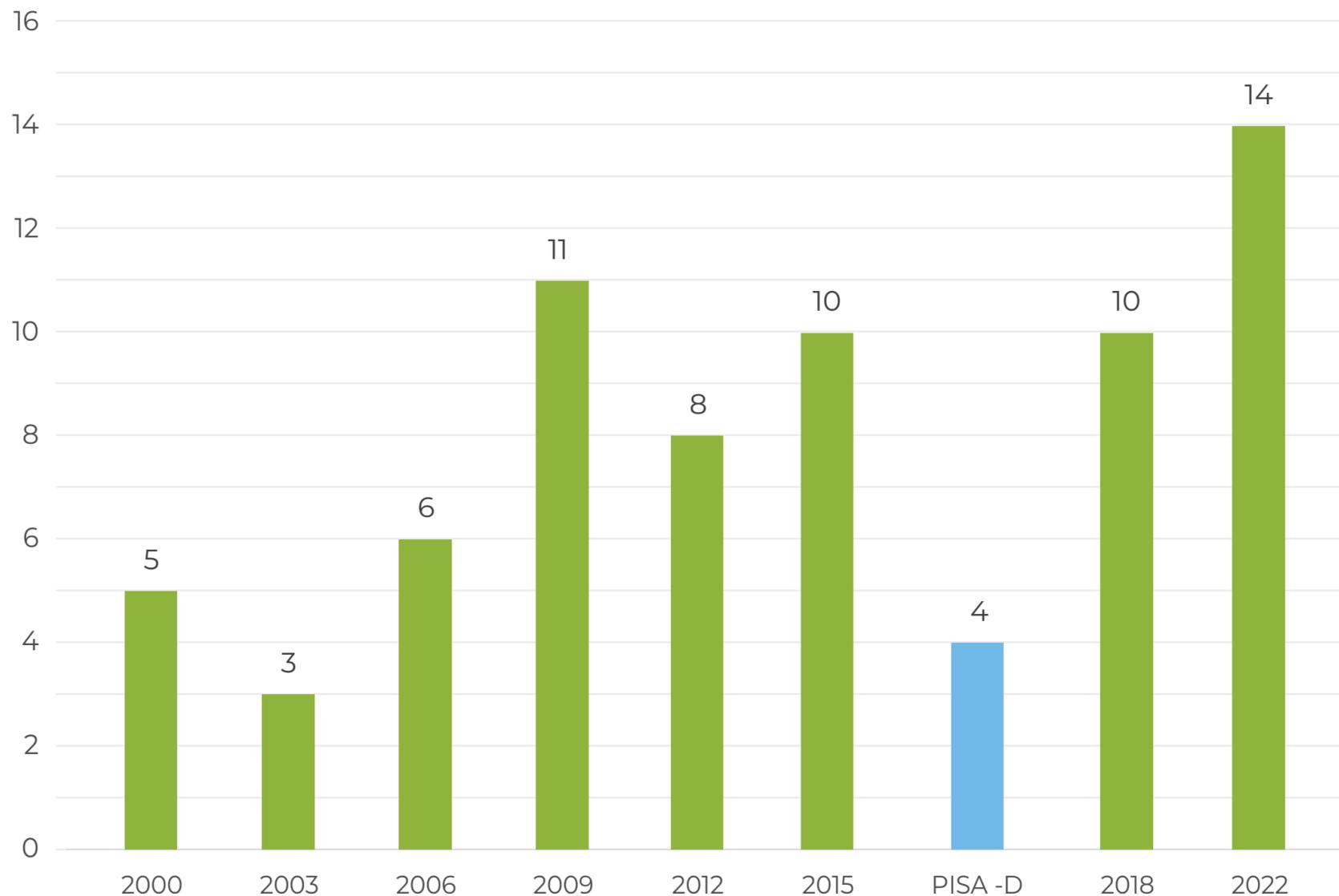
Evaluaciones internacionales

La creación del Programa Internacional para la Evaluación de Estudiantes (PISA) en 2000 trazó una agenda política común hacia la medición de la calidad educativa para fomentar políticas basadas en evidencia. La evaluación PISA se aplica a estudiantes de quince años que cursen al menos séptimo grado y mide si las y los estudiantes son capaces de reproducir lo que han aprendido en matemáticas, lectura y ciencia y su capacidad para aplicar esos conocimientos en entornos desconocidos, tanto dentro como fuera de la escuela (Arias Ortiz et al., 2024). PISA evalúa los conocimientos y habilidades en matemáticas, ciencia y lectura y se aplica cada tres años. Sin embargo, la octava ronda, que estaba prevista para

2021, se pospuso a 2022 debido a la pandemia del COVID-19.

La primera ronda de PISA (2000) contó con la participación de 43 países incluyendo a cinco latinoamericanos: Argentina, Brasil, Chile, México y Perú. La participación de países de la región ha ido creciendo de manera sostenida y en 2022 hubo una participación récord de catorce países de América Latina y el Caribe. Cabe destacar que desde la década de 2010 las evaluaciones internacionales como PISA ya no se limitan únicamente a países de ingresos medios y altos como en sus inicios, sino que comienzan a cubrir a más países de ingresos medios y bajos (Ramírez et al., 2018).

Figura 18: cantidad de países participantes en PISA



Fuente: elaboración propia sobre la base de datos de PISA, OCDE.

La evaluación PISA para el Desarrollo (PISA-D), llevada a cabo en 2017, merece una mención especial por su relevancia para la región. PISA-D surgió para adaptar la evaluación a los contextos de países de ingresos bajos y medios, introduciendo ítems diseñados para medir habilidades muy básicas. Participaron siete países, incluyendo a cuatro de la región: Ecuador, Guatemala, Honduras y Paraguay. El objetivo principal fue discriminar mejor y revelar diferencias en el desempeño de estudiantes que habrían quedado clasificados en el nivel “Por debajo de 1” en la versión regular de PISA (OCDE, 2018). Además, PISA-D brindó a los países con unidades de evaluación menos desarrolladas la oportunidad de fortalecer sus capacidades técnicas con miras a su eventual integración en el programa de evaluación global regular (Ferrer y Fiszbein, 2015). Tras su participación en PISA-D, Paraguay y Guatemala participaron de PISA 2022.

Muy pocos países de la región participan en otras evaluaciones internacionales. La Asociación Internacional para la Evaluación del Rendimiento Educativo²⁹ (IEA, por sus siglas en inglés) lleva a cabo evaluaciones a gran escala periódicamente desde 1995. Dentro de las evaluaciones de IEA se destacan la evaluación de desempeño de las y los estudiantes en matemáticas y ciencias (TIMSS, por sus siglas en inglés), comprensión lectora (PIRLS, por sus siglas en inglés), educación cívica y ciudadana (ICCS, por sus siglas en inglés) y competencia digital de las y los estudiantes (ICILS, por sus siglas en inglés). La participación de países de América Latina y el Caribe en estas cuatro pruebas es muy baja y menor aún en las últimas rondas. Chile se destaca como el único país de la región que ha participado en las cuatro evaluaciones internacionales mencionadas, aunque no ha participado en las últimas rondas de TIMSS e ICCS Colombia, Argentina y Brasil han participado de tres de las cuatro evaluaciones, aunque sólo Brasil se ha mantenido activo en las últimas rondas.

²⁹ Organización internacional de centros e instituciones nacionales de investigación, agencias de investigación gubernamentales, académicos y analistas que trabajan para investigar, comprender y mejorar la educación alrededor del mundo.

Tabla 1: participación de países de ALC en pruebas internacionales IEA

Países	PIRLS					TIMSS								ICCS			ICILS		
	2001	2006	2011	2016	2021	1995	1999	2003	2007	2011	2015	2019	2023	2009	2016	2022	2013	2018	2023
Chile				✓	✓		✓	✓		✓	✓			✓	✓		✓	✓	✓
Colombia	✓		✓			✓			✓					✓	✓	✓			
Brasil					✓	✓						✓	✓			✓			
Argentina	✓			✓				✓			✓						✓		
Trinidad y Tobago		✓	✓	✓															
Honduras			✓							✓									
Uruguay																		✓	✓
Belice	✓																		
Paraguay														✓					
El Salvador									✓										
México														✓	✓				
Perú															✓				
Guatemala														✓					
Rep. Dominicana														✓	✓				
ALC	3	1	3	3	2	2	1	2	2	2	2	1	1	6	5	2	2	2	2

Fuente: elaboración propia sobre la base de datos de la IEA.

¿Por qué son importantes estas evaluaciones para los sistemas nacionales de evaluación?

Las evaluaciones a gran escala regionales e internacionales, y especialmente ERCE y PISA, han contribuido significativamente al desarrollo y evolución de los sistemas educativos nacionales de la región. La participación en pruebas internacionales no sólo permite comparar los sistemas educativos entre países, sino que brinda además la oportunidad de compartir técnicas, estructuras organizativas y políticas que han demostrado ser eficientes y exitosas en otros contextos. La disponibilidad de información detallada y comparativa facilita la identificación de áreas específicas de alto rendimiento y destaca las áreas de mayor debilidad (Banco Mundial, 2011).

Efectivamente, la participación en pruebas internacionales y regionales ha sido una oportunidad clave para fortalecer la institucionalidad y las capacidades de los sistemas nacionales de evaluación educativa. Estas experiencias han impulsado mejoras tanto en el diseño como en la implementación de evaluaciones, permitiendo a los países avanzar en la consolidación de sus propias capacidades técnicas en este ámbito (UNESCO, 2014).

Además, la participación en evaluaciones internacionales ha sido determinante para la creación y la reforma de agencias y programas nacionales de evaluación (Tobin, Nugroho y Lietz, 2016). La participación en evaluaciones como PISA ha sido un factor clave para el fortalecimiento de las evaluaciones a gran escala en Brasil, contribuyendo a la convergencia técnica y política y solidificando el compromiso con el seguimiento de las políticas educativas (Fernandes y Gremaud, 2020). Concretamente, la participación en PISA en 2000 y 2003 impulsó reformas en el Sistema Nacional de Avaliação do Ensino Básico (SAEB), transformándolo en una evaluación censal que permite análisis más profundos de subgrupos y factores asociados con el rendimiento (Saracho Martínez, 2007).

LLECE se ha establecido como un referente para los países latinoamericanos en la evaluación educativa, proporcionando un marco de cooperación y apoyo técnico que ha permitido la formación y capacitación de los equipos responsables de los sistemas nacionales de medición y evaluación (Vanni y Valenzuela, 2020). La colaboración horizontal promovida por LLECE ha sido crucial. Balarin (2021) subraya que los países latinoamericanos desempeñaron un papel central en la definición

de los dominios de aprendizaje de ERCE y en el desarrollo de ítems de prueba, contribuyendo directamente al fortalecimiento de los sistemas de evaluación nacionales.

El proceso de preparación y aplicación de las pruebas PISA también ha servido como un mecanismo para desarrollar y fortalecer capacidades técnicas y metodológicas a nivel local, facilitando el intercambio de conocimientos y la adquisición de metodologías de medición más avanzadas (Ferrer y Fiszbein, 2015). Asimismo, ha permitido la exposición a ideas innovadoras y mejores prácticas en el campo de la evaluación a gran escala (Rivas, 2015). El trabajo conjunto con especialistas en las pruebas PISA ha permitido integrar nuevas capacidades en las agencias locales, especialmente en áreas como la teoría de respuesta al ítem, la definición de niveles de desempeño, el establecimiento de líneas de corte y la optimización de la confiabilidad en la corrección de preguntas abiertas (Ferrer, 2009). Además, se han incluido instrumentos para medir factores contextuales que ayudan a explicar los rendimientos de las y los estudiantes (Altinok et al., 2018).

Aunque los países de la región no tienen un papel protagónico en el diseño y desarrollo de las pruebas PISA, las unidades nacionales de evaluación han encontrado en estas pruebas una plataforma valiosa para el desarrollo y fortalecimiento de sus capacidades técnicas y metodológicas. Efectivamente, se constata un significativo proceso de aprendizaje en diversos aspectos técnicos de la evaluación a gran escala, como el diseño de pruebas, el trabajo con preguntas de respuesta abierta, la creación de instrumentos que permitan comparaciones temporales y el control de calidad en los procesos de muestreo, aplicación y corrección de pruebas (Ferrer y Fiszbein, 2015).

La participación en pruebas internacionales también tiene importancia en la promoción de políticas educativas nacionales. Las pruebas pueden apalancar prioridades de políticas educativas y, en algunos casos, crear un entorno favorable para reformas educativas, desta-

cándose como un elemento central al analizar su influencia internacional. A nivel global, la literatura revela que el impacto de las pruebas en el apalancamiento puede ir en ambas direcciones. Por ejemplo, los resultados de PISA 2000 en Alemania provocaron un shock en la opinión pública y el ámbito político, abriendo una ventana para reformas (Heyneman y Lee, 2014). Sin embargo, en Inglaterra los mismos resultados sirvieron para justificar y legitimar las políticas ya existentes (Cox y Meckes, 2016). En países de ingresos medios y bajos, el impacto de las evaluaciones internacionales en la definición de la agenda se observa frecuentemente a través de comparaciones internacionales que inician debates sobre una calidad educativa inferior a la esperada (Tobin, Nugroho y Lietz, 2016). En América Latina, la participación en PISA ha puesto de relieve el tema de la equidad educativa al visibilizar las brechas de logro entre diferentes segmentos de la población (Rivas, 2015). Además, en la región las evaluaciones internacionales están más asociadas con el apalancamiento de políticas que otros tipos de evaluación, como las regionales o nacionales (Tobin, Nugroho y Lietz, 2016). Por ejemplo, el bajo desempeño de Brasil en PISA 2000 y 2003 ayudó a catalizar y apoyar el movimiento hacia la evaluación y reforma educativa que había comenzado en la década de 1990 (Saracho Martínez, 2007). En Perú los resultados de las pruebas PISA influyeron sobre la política educativa de la primera década del siglo en tanto la conmoción causada por los resultados del país en PISA ayudó a generar un ambiente propicio para la discusión sobre las necesidades de reforma educativa (Cueto, 2007).

Algunos países de América Latina ofrecen ejemplos interesantes en los que la participación en evaluaciones regionales e internacionales a gran escala ha influido en reformas concretas de política educativa. La literatura sugiere que, a partir de su participación en programas de evaluación internacionales, los sistemas educativos suelen incorporar nuevos conocimientos y habilidades en sus cu-

rrículos, lo que a menudo se traduce en reformas curriculares y en la definición o ajuste de estándares de desempeño (Cox y Meckes, 2016). En 2009 Chile alineó sus currículos de matemáticas y ciencias con los estándares de TIMSS y utilizó el marco del Estudio de Educación Cívica (CIVED) de la IEA para reformar la educación en ciudadanía, integrándola en todo el currículo y enfatizando habilidades y conocimientos cívicos (Cariola et al., 2011). Ferrer y Fiszbein (2015) subraya que, a partir de PISA, las competencias de comprensión lectora se han estandarizado en tres procesos cognitivos distintos pero complementarios: localizar información en un texto, hacer inferencias y evaluar la calidad del texto para transmitir el significado. Antes de PISA existía una gran diversidad en la definición de “lectura”, con variaciones según las pruebas o currículos implementados. En otros casos, la participación en este tipo de evaluaciones tiene incidencia en reformas relativas a la formación docente. Los resultados de PISA informaron el desarrollo de un programa federal de formación profesional, la Universidad Abierta de Brasil (UAB), que proporciona programas de educación a distancia y capacitación para maestros y líderes escolares en diversas áreas (Tobin, Nugroho y Lietz, 2016).

Ferrer y Fiszbein (2015) reflexiona sobre el valor de PISA para los países de la región. Una de las objeciones más comunes es la percepción de que no tiene sentido invertir tiempo y recursos en una evaluación en la que los países de la región suelen ocupar los últimos lugares en el ranking. Sin embargo, los autores argumentan que esto ignora el valor que significa que PISA proporcione cada tres años una medida regular con instrumentos confiables. Además, los resultados pueden generar debates nacionales de interés, poniendo en el centro a la discusión de la calidad educativa, lo que puede resultar en mayor apoyo para implementar reformas. Como hemos dicho, estas pruebas pueden fortalecer las capacidades técnicas de los equipos nacionales de

evaluación, actuar como una herramienta de presión para apalancar prioridades políticas y servir como catalizadores de reformas y mejoras educativas. Pero es esencial reconocer que no son adecuadas para diseñar políticas y programas específicos a nivel de escuela o aula ni para abordar necesidades de sectores vulnerables o rezagados. Para ello es necesario contar con evaluaciones nacionales y locales que ofrezcan un diagnóstico más detallado y contextualizado de las necesidades y desafíos de cada escuela y comunidad educativa. Por ello es clave la complementariedad entre evaluaciones internacionales y nacionales.

La participación en evaluaciones internacionales y regionales adquiere particular relevancia para el seguimiento de las metas establecidas en el Objetivo de Desarrollo Sostenible 4. Los objetivos relacionados con la calidad de los aprendizajes en habilidades fundamentales (matemáticas y lectura) se enmarcan en la meta 4.1, que establece: “De aquí a 2030, asegurar que todas las niñas y todos los niños terminen la enseñanza primaria y secundaria, que ha de ser gratuita, equitativa y de calidad y producir resultados de aprendizaje pertinentes y efectivos” (Asamblea General de las Naciones Unidas, 2015). Para garantizar el adecuado monitoreo de las mejoras en la calidad de la educación básica en habilidades fundamentales se ha definido un grupo de indicadores (4.1.1) que miden la proporción de niños, niñas y adolescentes que han alcanzado al menos un nivel mínimo de competencia en lectura y matemáticas en tres momentos del ciclo escolar: (a) al final del segundo o tercer grado, (b) al final de la enseñanza primaria y (c) al final de la enseñanza secundaria baja.

El seguimiento de los indicadores de la meta 4.1 enfrenta el desafío de la escasez de datos precisos y comparables para evaluar los resultados de aprendizaje y su evolución a nivel global a lo largo del tiempo (Montoya, 2023). A pesar de la adopción de medidas recientes para obtener datos de aprendizaje confiables y comparables (ver Box 2), más de 60 por cien-

to de los países a nivel global no cuenta con datos sobre calidad de los aprendizajes. De los veintiséis países de América Latina y el Caribe miembros del BID, 16 países han realizado durante el período 2016-2022 un seguimiento de los niveles de competencia en matemáticas y lectura al final del segundo o tercer grado y al final de la enseñanza secundaria baja; y 17 países lo han hecho al final de la primaria, todos con base en su participación en pruebas regionales y/o internacionales.

ERCE se ha consolidado como la principal herramienta de monitoreo de aprendizajes en la agenda 2030 para América Latina y el Caribe, permitiendo el seguimiento en 16 países de la región³⁰. ERCE monitorea los indicadores de lectura y matemáticas al final del segundo o tercer grado y al final de la primaria. La participación de Trinidad y Tobago en PIRLS permite elevar el número de países que reportan este indicador a 17 para lectura al final de la primaria (considerando el período 2016-2022). Por otro lado, PISA y PISA-D son la principal

fuerza de información para el seguimiento de ambas asignaturas al final de la secundaria baja, permitiendo el monitoreo en 14 países de la región para el año 2022 y en dieciséis países si se considera el período 2017-2022. A excepción de Trinidad y Tobago y Jamaica, los países del Caribe no participan actualmente en evaluaciones regionales o internacionales que cumplan los criterios UIS (Box 2), lo que dificulta el seguimiento y monitoreo de los aprendizajes en línea con la meta ODS 4.1.

30 En Bolivia se reporta para 2017 con datos del Diagnóstico Nacional realizado con la aplicación de los instrumentos del TERCE aunque no son estrictamente comparables.

Box 2:

Evaluaciones de aprendizaje y seguimiento del Objetivo de Desarrollo Sostenible 4

Ante la falta de datos, la Alianza Mundial para el Seguimiento del Aprendizaje (GAML), con el apoyo técnico del Australian Council for Educational Research (ACER) y el Global Education Monitoring Report (GEMR), ha definido estándares globales de Nivel Mínimo de Competencia (Minimum Proficiency Level, MPL) y su correspondencia con los resultados de una serie de evaluaciones internacionales y regionales. Esto permite usar diversas fuentes de información para el seguimiento de la calidad de los aprendizajes con el objetivo de promover mayores niveles

de cobertura (UNESCO Institute for Statistics, 2024). Los programas de evaluaciones internacionales que se utilizan para informar sobre el grupo de indicadores 4.1.1 son: el Programa para la Evaluación Internacional de Estudiantes (PISA), el Estudio Internacional de Lectura y Alfabetización (PIRLS) y el Estudio Internacional de Tendencias en Matemáticas y Ciencias (TIMSS). El protocolo considera también un grupo de cinco evaluaciones regionales aplicadas a nivel global, incluyendo a Estudios Regionales Comparativos y Explicativos (ERCE) como prueba regional para América Latina y el Caribe.

	Grado 2° o 3°		Final de la primaria		Final de la secundaria baja	
	Lectura	Matemáticas	Lectura	Matemática	Lectura	Matemáticas
PIRLS (I)	✓		✓			
TIMSS (I)		✓		✓		✓
PISA/PISA-D (I)					✓	✓
SEA-PLM (R-Asia Sudeste)			✓	✓		
ERCE (R-ALC)	✓	✓	✓	✓		
PASEC (R-África)	✓	✓	✓	✓		
SACMEQ (R-África)			✓	✓		
PILNA (R-Pacífico)	✓	✓	✓	✓		

Fuente: GAML (2023). I=evaluación internacional, R=evaluación regional.

Por su parte, el potencial uso de evaluaciones nacionales de aprendizaje como fuente de reporte a los indicadores de ODS 4.1.1 aún se encuentra en fase piloto y está sujeto al cumplimiento de los seis criterios establecidos por el GAML (UNESCO Institute for Statistics, 2024):

1. la evaluación utiliza la teoría de respuesta al ítem (IRT);
2. los resultados están disponibles como porcentaje de estudiantes por niveles de competencia;
3. un descriptor de nivel de competencia está alineado con el nivel mínimo de competencia (MPL);
4. el contenido cubre suficientemente el marco de contenido global en el dominio relevante, ya sea en **lectura** y/o **matemáticas**. La herramienta de alineación de contenido permite este mapeo y la orientación se puede encontrar en el informe **(CAT)**;
5. cumple con el nivel mínimo de buenas prácticas, la herramienta de alineación de procedimientos permite a los países evaluar el nivel de cumplimiento con un conjunto de buenas prácticas y, lo más importante, alcanzan un nivel mínimo **(PAT)**; y
6. se añaden características a los puntos de datos (nombre de la evaluación nacional, nivel mínimo de competencia y grado).

Finalmente, el GAML también ha desarrollado la evaluación para el Nivel Mínimo de Competencia **(AMPL)** que permite determinar si un estudiante ha alcanzado el nivel mínimo de competencia en matemáticas o lectura para su nivel educativo específico. Estas herramientas producen resultados comparables entre países y sirven para informar sobre los indicadores 4.1.1. a y b. Las evaluaciones forman parte de un conjunto de herramientas metodológicas accesibles para medir aprendizajes, desarrollar capacidades para generar, analizar e informar periódicamente datos de aprendizaje, y monitorear el progreso hacia el ODS 4.

II.2 Las evaluaciones de Aprendizaje nacionales en la región: evolución y estado actual

Las evaluaciones nacionales de aprendizaje en América Latina³¹ evolucionaron en tres grandes fases. Una primera etapa durante la década de 1990 de creación de capacidades evaluativas; una segunda en la primera década del siglo donde en términos generales se ampliaron los sistemas; y una tercera desde la década de 2010 hasta la disrupción de la pandemia donde se vio una consolidación y la revisión de objetivos y conceptos evaluativos (Galas, Gutiérrez y Hamilton, 2022). Si bien estas fases enmarcan las grandes discusiones y transformaciones técnicas en la medición de aprendizajes en la región, existe una amplia heterogeneidad en las características con las que han evolucionado tanto los sistemas de evaluación como las pruebas de aprendizaje nacionales a gran escala. A continuación, profundizamos esta mirada haciendo énfasis en las características clave de las pruebas de aprendizaje en cuanto a alcance, calidad y gobernanza.

Creación de capacidades evaluativas (1990-2000)

La primera fase se caracteriza por la creación de capacidades evaluativas. Las primeras evaluaciones fueron impulsadas por cooperaciones internacionales, relevadas por universidades, ministerios de educación e instituciones evaluativas. Se medían los aprendizajes en matemáticas y lengua bajo un enfoque normativo y psicométrico de la teoría clásica del test.

En la mayoría de los países de la región la experiencia en evaluación educativa a gran escala se fue adquiriendo lentamente a través de la administración de pruebas con propósitos específicos como regular la ad-

misión a la educación superior (Colombia), medir el impacto de un proyecto de mejora educativa financiado por una organización internacional (Guatemala, Ecuador y Honduras) o regular el sector privado y proporcionar información a las familias para mejorar sus decisiones en un contexto de mayor libertad de elección escolar (Chile). En la década de 1990 varios países adoptaron la evaluación a través de pruebas estandarizadas como parte de los programas de reforma educativa más o menos comprensivos de la década. Como se verá más adelante, estas evaluaciones se llevaron a cabo dentro de marcos institucionales y organizacionales creados específicamente para fines de evaluación (Ferrer, 2006). En casos excepcionales, como el de Chile, el enfoque de las pruebas cambió rápidamente hacia el uso de sus resultados para focalizar mejor los recursos y las intervenciones en las escuelas de bajo desempeño (Elacqua y Fabrega, 2004). En 1997 los resultados de las evaluaciones fueron decisivos para que el Congreso chileno aprobara una ampliación del horario escolar anual de 40 por ciento con el objetivo de abordar el déficit de calidad educativa detectado en 1994, cuando casi el 40 por ciento de las y los estudiantes de cuarto grado no comprendían los textos que leían (Schiefelbein y Schiefelbein, 2003). En Uruguay, los resultados de la primera evaluación nacional en 1996 demostraron la relación entre el contexto sociocultural de las y los estudiantes y sus resultados de aprendizaje, destacando la necesidad de integrar estos factores en las reformas educativas (Benveniste, 2000).

En esta etapa el financiamiento y la institucionalidad de los procesos de evaluación eran diversos y cambiantes. La mayoría de los países de la región comenzaron el proceso de evaluación de los aprendizajes utilizando distintas fuentes de financiamiento, ya sea internacional³² o nacional, y las incipientes responsabilidades de evaluación fueron atribuidas a departamentos especializados en

31 Con la temprana creación de LLECE apareció un intenso intercambio y colaboración dentro de América Latina que no se verificó en el mismo grado en el Caribe. Por ello, aquí nos enfocamos en la trayectoria de los sistemas de evaluación nacional a gran escala de los 18 países de América Latina sin incluir a los del Caribe.

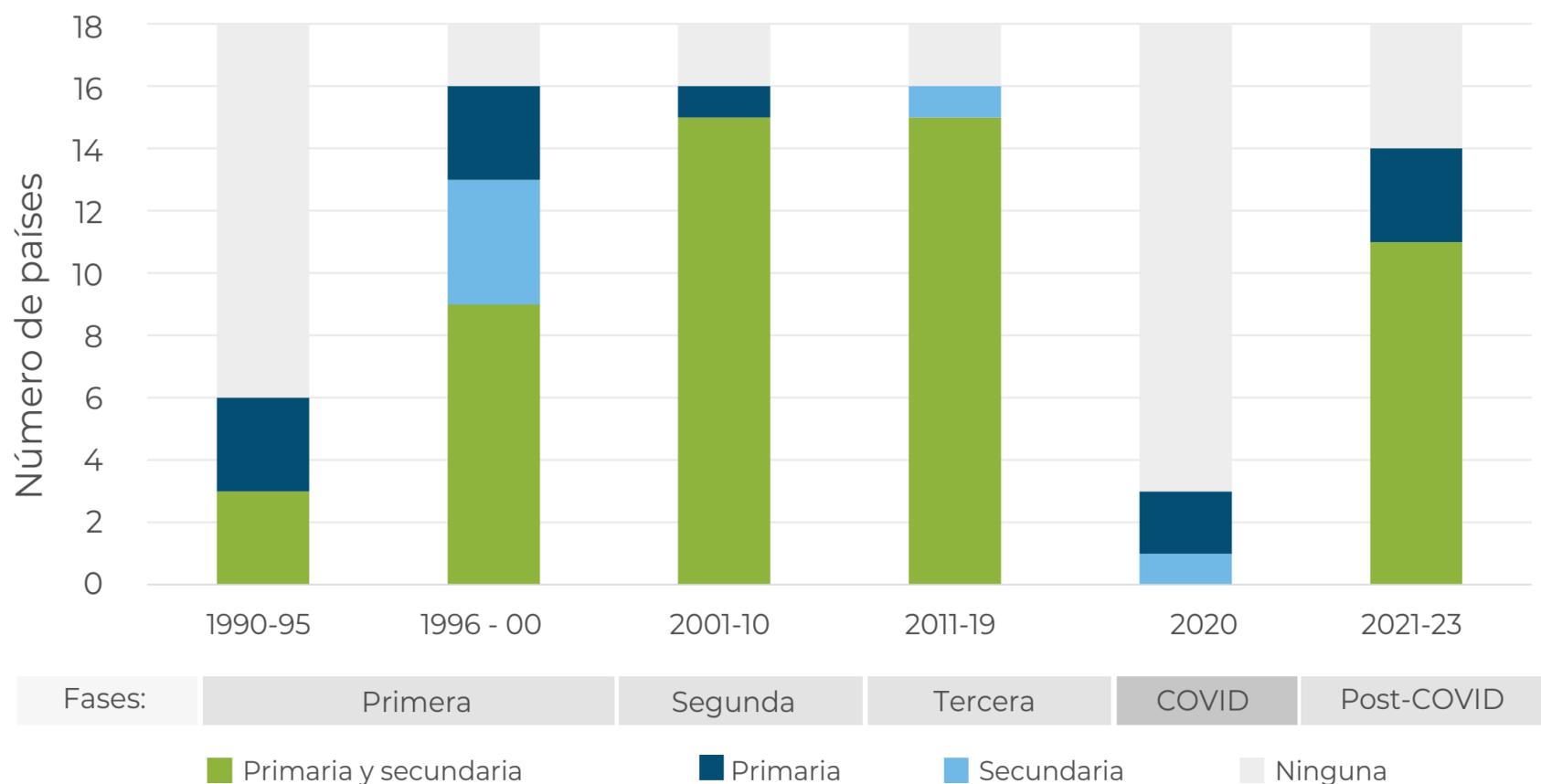
32 Las primeras pruebas nacionales de aprendizaje en países como Guatemala, Ecuador, Honduras y Uruguay fueron financiadas y promovidas por organismos internacionales.

los ministerios de educación (Argentina, Bolivia, Ecuador, El Salvador, México, Paraguay, Perú, República Dominicana, Uruguay y Venezuela); instituciones existentes que ampliaban sus funciones (el INEP en Brasil y el ICFES en Colombia); y universidades, partiendo de su experiencia técnica y con alianzas con los ministerios de educación (Chile, Costa Rica, Guatemala y Honduras).

A nivel temático las evaluaciones se enfocaban en las áreas de matemáticas y lengua y a nivel metodológico el enfoque de las pruebas era normativo. Las evaluaciones en este período se focalizaron en la medición del rendimiento académico en las áreas de lenguaje y matemáticas; sólo Chile, Brasil y Argentina medían ciencias naturales y sociales y Uruguay actitudes ciudadanas. En la mayoría de los países se aplicaba la metodología convencional de pruebas y un enfoque normativo donde los resultados se interpretan con

respecto a un grupo particular usualmente referido como el grupo estándar o normal, que resumían los resultados en términos de puntajes o cantidad de respuestas correctas, conocido como teoría clásica del test (TCT). Sólo Brasil, Chile y México incluyeron tempranamente un enfoque de teoría de respuesta al ítem (TRI), utilizando un enfoque de criterios dentro de las pruebas; es decir, que los resultados eran interpretados con referencia a una habilidad determinada o a lo que el estudiante sabe hacer (Harris, 1991). La teoría de respuesta al ítem ofrece una descripción más detallada de lo que las y los estudiantes saben y pueden hacer, permitiendo evaluar el nivel de conocimiento de manera más precisa y ofreciendo un potencial mayor para intervenciones pedagógicas en comparación con los informes basados únicamente en el porcentaje medio de respuestas correctas (Ferrer y Fiszbein, 2015).

Figura 19: países con pruebas de aprendizaje a gran a escala por nivel educativo, 18 países (1990-2023)



Fuente: elaboración propia sobre la base de la revisión de la información de cada sistema.

En este período los países también tomaron decisiones sobre el nivel educativo a evaluar y la cobertura de sus evaluaciones (muestra o censal). Hasta la primera mitad de la década de 1990, sólo un tercio de los países de la región contaba con alguna prueba de aprendizaje nacional a gran escala (Figura 19). En educación primaria, sólo tres países realizaban pruebas (Argentina, Brasil y Chile), evaluando desempeños en matemáticas y lenguaje. Chile aplicaba pruebas censales para primaria en las áreas de matemáticas y lengua y medía de forma muestral las áreas de ciencias naturales y sociales. En educación secundaria, sólo Colombia y Costa Rica realizaban pruebas nacionales censales para las y los estudiantes de último año de la educación media, y Argentina comenzó a evaluar de forma muestral a las y los estudiantes de primer y último año de educación secundaria en matemáticas y lengua.

Hacia finales de la década, 16 de los 18 países analizados habían realizado al menos una primera evaluación en primaria (Bolivia, Honduras, Guatemala y México), secundaria (Colombia, República Dominicana y El Salvador) o en ambos niveles (Argentina, Brasil, Chile, Costa Rica, Ecuador, Perú, Paraguay, Uruguay y Venezuela). Las pruebas en primaria fueron, en general, aplicadas a estudiantes del último año y en grados intermedios como tercero y cuarto, mientras que las pruebas en secundaria eran aplicadas principalmente a estudiantes en el último año de secundaria baja (Figura 20). Las pruebas carecían de una repercusión y aplicabilidad significativa en política educativa, llegando incluso a existir resistencia a la divulgación de sus resultados³³. Sin embargo, en el sistema chileno desde muy temprano se comenzaron a utilizar las pruebas censales para crear rankings de acuerdo con el puntaje de las escuelas. El primer ordenamiento que comenzó a publicar el Sistema de Medición de la Calidad de la Educación (SIMCE) carecía de un control por características socioeconómicas de las instituciones, suscitando interpreta-

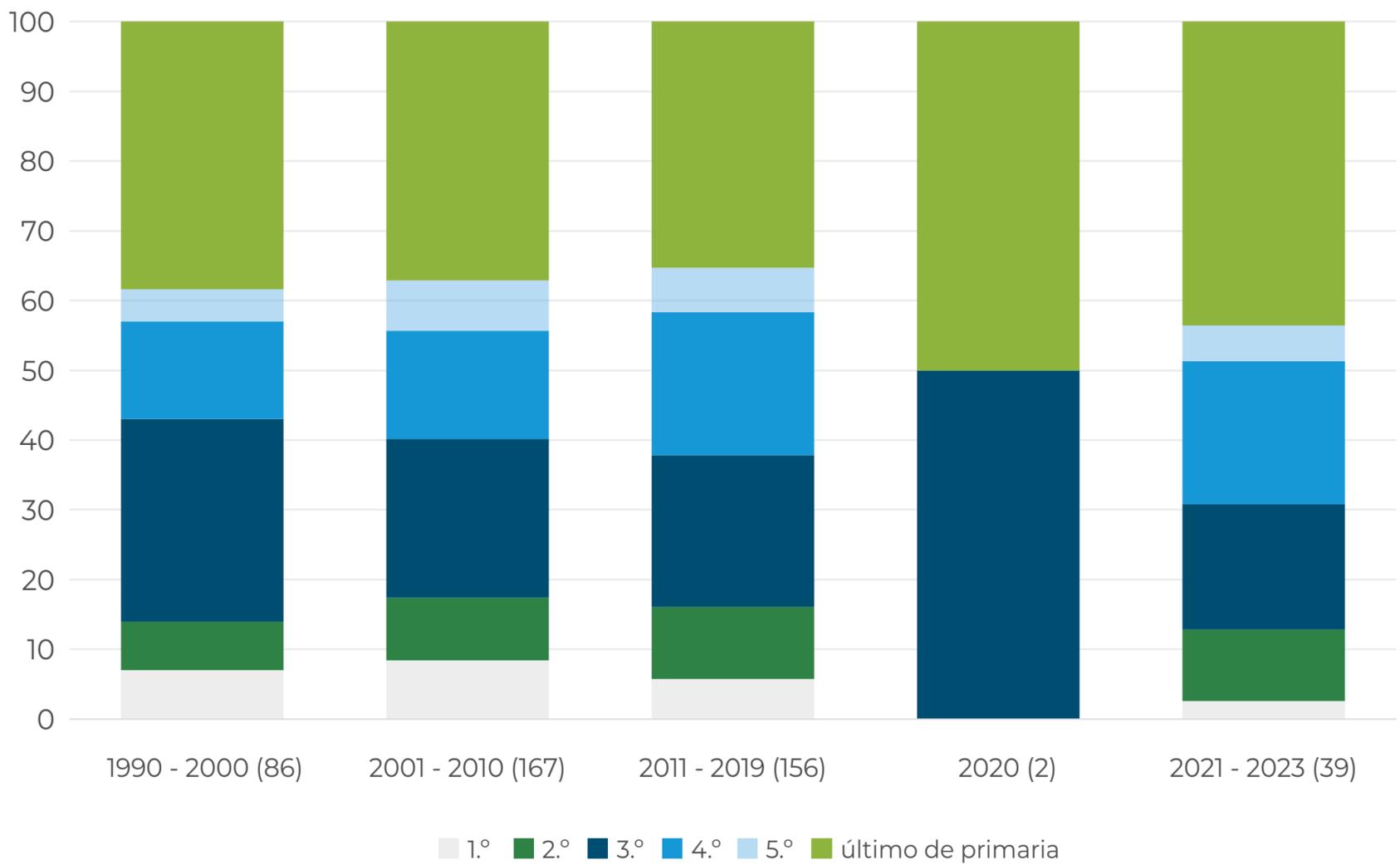
ciones erróneas y críticas por la poca comparabilidad de los resultados³⁴. Por otro lado, las pruebas censales en el último año de secundaria en República Dominicana, El Salvador y Costa Rica se caracterizaban por tener altas implicancias, buscando establecer un estándar de aprobación del nivel, mientras que en Colombia eran consideradas un requisito para ingresar a una institución de educación superior. Las pruebas muestrales en los niveles intermedios o últimos de secundaria tuvieron, en cambio, objetivos diversos tendientes hacia objetivos formativos. En Argentina, Brasil y Ecuador, por ejemplo, se publicaban informes con recomendaciones metodológicas para los y las docentes; estos discutían algunos de los ítems de las pruebas y explicaban las dificultades que los y las estudiantes habían encontrado en los exámenes. Sin embargo, los resultados de las evaluaciones no tenían lugar en el debate público ni en la agenda del sistema educativo (Ferrer, 2006).

33 En 2000, México y Perú censuraron la publicación de los resultados obtenidos en pruebas internacionales y nacionales (Ferrer y Fiszbein, 2015).

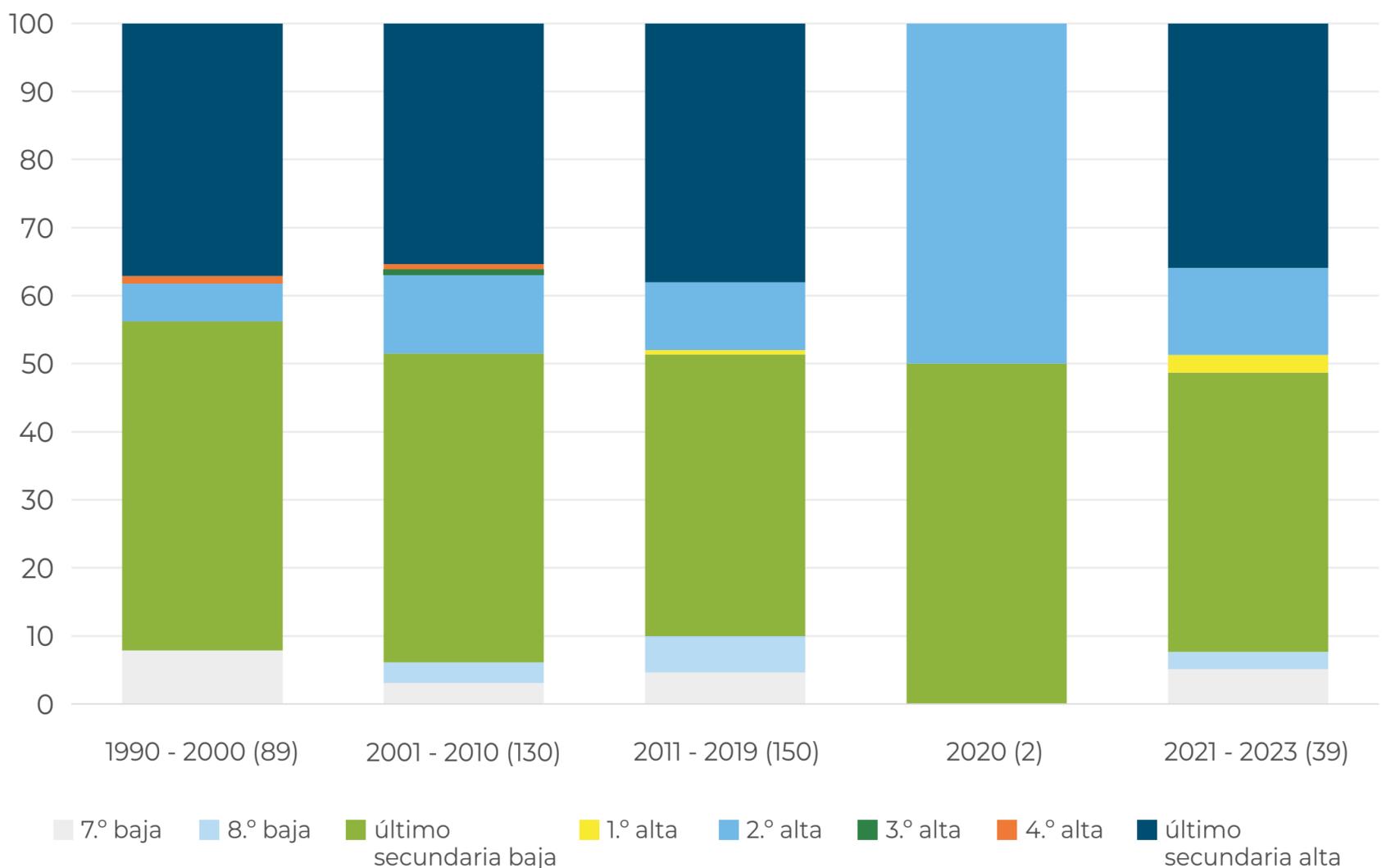
34 En esta época la prensa chilena elaboraba rankings de escuelas utilizando la información publicada por el gobierno para cada establecimiento. Estos rankings no consideraban factores contextuales y comparaban directamente todas las escuelas, tanto públicas como privadas. No fue hasta 2008, con la implementación de la Ley SEP, que el gobierno comenzó a publicar rankings más robustos, ajustados por factores externos.

Figura 20: grados evaluados en pruebas de aprendizaje a gran escala por nivel educativo, 18 países (1990-2023)

Primaria – CINE 1³⁵



Secundaria – CINE 2 y 3



Fuente: elaboración propia sobre la base de la revisión de la información de cada sistema.

Nota: entre paréntesis se muestra el número de pruebas para cada año.

35 En términos de la clasificación del logro educativo, la Clasificación Internacional Normalizada de la Educación (CINE) clasifica de forma homogénea los niveles educativos de primaria (CINE 1), educación secundaria baja (CINE 2) y educación secundaria alta (CINE 3). Para mayor detalle ver UNESCO-UIS (2013).

Cambios institucionales y ampliación de los sistemas de evaluación educativa (2001-2010)

La primera década del siglo XXI se caracterizó por cambios en el marco institucional de las evaluaciones nacionales, una tendencia hacia la recolección de pruebas en primaria y secundaria que hacían uso de formularios de contexto, presentaban resultados normativos y un mayor avance hacia enfoques de teoría de respuesta al ítem. Las evaluaciones tenían mayor difusión y uso.

El marco institucional de los sistemas de evaluación de la educación tuvo varios cambios en esta década.

En Costa Rica y Chile los ministerios de educación asumieron la responsabilidad de las evaluaciones que antes compartían con las universidades. En Panamá, el ministerio designó a la Dirección Nacional de Evaluación Educativa (DNEE) para el diseño y aplicación de evaluaciones para la generación de información para la toma de decisiones. En México se creó en 2002 el Instituto Nacional para la Evaluación de la Educación (INEE) como organismo público dependiente, y en 2012 obtuvo autonomía por ley en el marco de la reforma educativa. Argentina creó en el año 2000 el Instituto de Calidad Educativa, una agencia semiautónoma del Ministerio de Educación que poco después fue reemplazada por la Dirección Nacional de Información y Evaluación de la Calidad Educativa (DiNIECE), dependiente del ministerio. Los análisis de la época indican que los arreglos institucionales eran más estables cuando las responsabilidades se gestionaban fuera de la estructura de los ministerios de educación a través de institutos con mayor autonomía administrativa y técnica (Ferrer, 2006). En la siguiente sección retomaremos esta discusión incluyendo otros casos en la región y la evolución de los arreglos institucionales.

En cuanto al relevamiento, se mantuvo el número de países con al menos una prueba y aumentó la proporción de países que recogían información para ambos niveles.

16 de los 18 países de la región realizaron alguna evaluación. Colombia, Costa Rica y Panamá implementaron pruebas a gran escala a nivel primaria para el último grado y grados intermedios (en el caso de Colombia) con fines diagnósticos.

Adicionalmente, se amplió la cobertura temática y la información de contexto recogida en las pruebas.

Las evaluaciones en este período siguieron focalizadas en las áreas de matemáticas y lengua y más países midieron a nivel primaria los resultados en ciencias naturales y sociales (Costa Rica, Ecuador, Panamá, República Dominicana, Colombia y México). México incluyó educación cívica dentro de sus evaluaciones y Uruguay dejó de medir actitudes ciudadanas e incluyó ciencias naturales. A comienzos de esta década también comenzaron a utilizarse ampliamente cuestionarios de contexto para mejorar la interpretación y disseminación de los resultados de aprendizaje. Con la paulatina consolidación de los sistemas de evaluación, hubo un creciente interés por ahondar en la relación que factores individuales, institucionales y docentes tenían en los resultados de aprendizaje. Aunque Uruguay, Brasil, Argentina, Perú, Chile y Colombia ya recogían formularios complementarios dirigidos a distintos tipos de actores educativos (padres de familia, docentes y directores), desde esta década Chile y Colombia comenzaron a integrar al análisis de los resultados índices socioeconómicos de las y los estudiantes y escuelas y el análisis de factores asociados que permitían realizar comparaciones más “justas” de los resultados y, hasta cierto punto, medir el valor agregado de la escuela dado que los resultados podían tener en cuenta las diferencias en el entorno de las y los estudiantes (ICFES, 2017; Rodríguez-Garcés, Padilla-Fuentes y Suazo Ruiz, 2020).

A nivel metodológico se inició una actualización de las teorías psicométricas utilizadas en las evaluaciones.

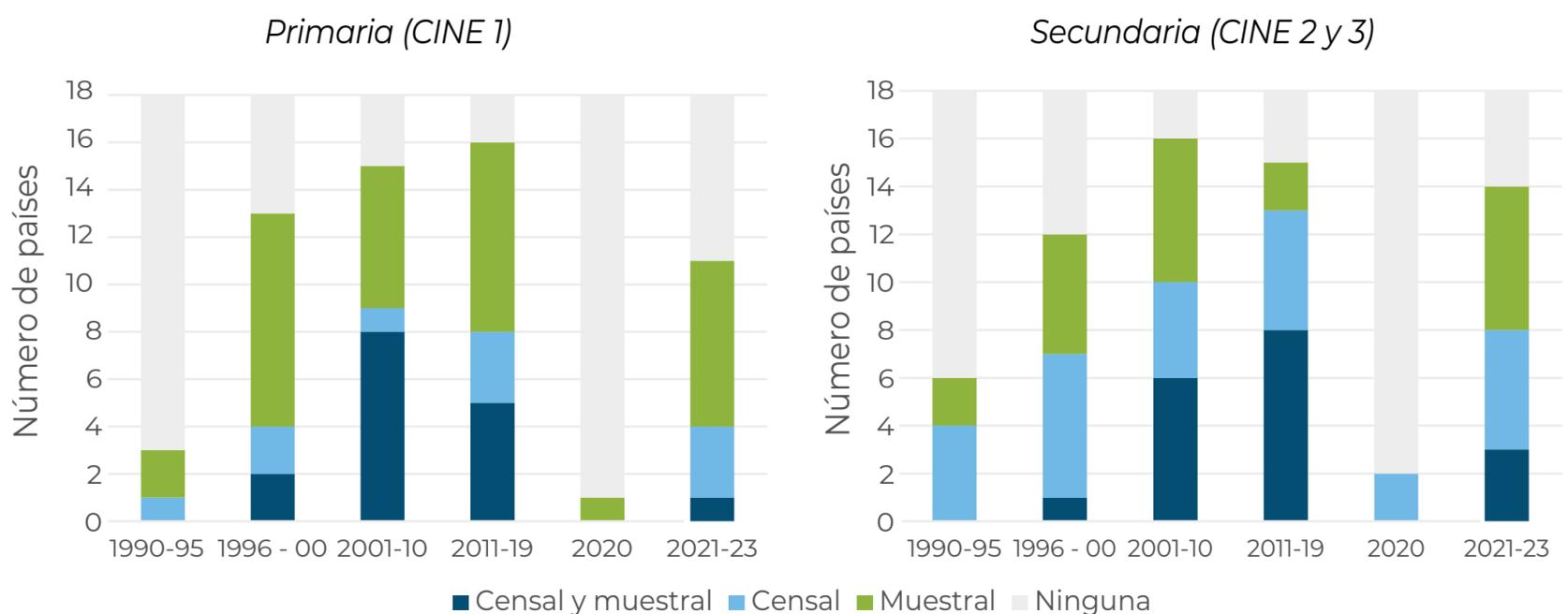
De hecho, hacia 2008 la mayoría de los países en la región, incluyendo a Bolivia, Argentina, Brasil, Chile, Guatemala y Ecuador, reportaron resultados por niveles de logro, en vez de puntajes (Galas, Gutiérrez y Hamilton, 2022) y 11 de los 18 países de Latinoamérica habían implementado en alguna de sus evaluaciones aproximaciones de la teoría de respuesta al ítem.

También se fortalecieron las prácticas de publicación de resultados y, conforme se consolidaban los aspectos técnicos, la recurrencia y tipos de pruebas, los países comenzaron a hacer mayor énfasis en el uso de su información.

Una de las primeras utilidades de los resultados y disseminación de las evaluaciones en la región fue la reflexión sobre las metas y estándares de aprendizaje a evaluar. De acuerdo con Arregui (2008), al comenzar la medición de los aprendizajes a gran escala los países de la región no tenían ideas claras sobre el uso que darían a los resultados. Una vez que comenzó el relevamiento de la información, en un proceso de ingeniería inversa, los países comenzaron a asignarles distintos usos. En esta década, teniendo en cuenta los avances técnicos de las pruebas, en Colombia, Perú y Brasil se crearon estándares educativos medibles en búsqueda del

diseño e implementación de políticas educativas basadas en sus resultados, se asignaron recursos para cerrar brechas de aprendizaje, se estimularon cambios pedagógicos y se promovió la responsabilización y la introducción de incentivos a varios actores e instituciones educativas mediante la asignación de implicancias específicas de los resultados de las pruebas a distintos actores (Galas, Gutiérrez y Hamilton, 2022; Ferrer y Fiszbein, 2015). Otro paso importante que tomó mayor relevancia en esta época en Colombia y Brasil, y sobre lo que se profundizará en la siguiente, fue el de poner los resultados de las evaluaciones a disposición de investigadores. En 1999 se consolidó en Brasil el Comité de los Productores de Información Educativa (COMEP) para promover la articulación entre productores de información en el área educacional y constituir un sistema integrado de difusión de información especializada. Sin embargo, la difusión y uso de los resultados aún eran debilidades clave en los sistemas de evaluación de aprendizaje de la región, con una baja participación de la sociedad civil, lo que no fomentaba la demanda de mayores niveles de calidad en la educación (Ferrer, 2006).

Figura 21 . países con pruebas nacionales de aprendizaje a gran a escala por tipo de prueba y nivel educativo, 18 países (1990-2023)



Fuente: elaboración propia sobre la base de la revisión de la información de cada sistema.

En esta línea, la elección entre un enfoque censal y muestral comenzó a depender de factores como el uso previsto de los resultados y el presupuesto disponible. En los sistemas donde las pruebas eran parte de un sistema de rendición de cuentas escolar para proporcionar retroalimentación formativa a las escuelas en forma de boletines informativos, o donde se buscaba identificar las escuelas de bajo desempeño y asignar recursos o evaluar políticas educativas, las evaluaciones censales fueron esenciales. En sistemas donde el objetivo era el de monitorear el rendimiento general del sistema educativo a lo largo del tiempo o comprender los factores contextuales que afectan el aprendizaje, fueron consideradas más eficientes las evaluaciones muestrales. Ambos enfoques pueden utilizarse para informar la mejora continua y la política educativa y práctica. Seleccionar una muestra de estudiantes para la evaluación resulta menos costoso que evaluar a todos los alumnos, pero limita la capacidad de generar informes detallados a nivel escolar, lo que restringe el uso de las evaluaciones para mejorar las prácticas de enseñanza en todas las escuelas. Algunos países, como Chile y Brasil, combinan enfoques basados en censos y muestras en el diseño de sus evaluaciones nacionales a gran escala para satisfacer de manera más efectiva las necesidades de información de las partes interesadas. La combinación ayuda a reducir los costos generales mientras proporciona una imagen detallada del sistema educativo (Clarke y Lune-Bazaldua, 2021). En la década entre el 2001-2010, de los dieciocho países analizados, en el nivel primario seis recogieron sólo evaluaciones muestrales, ocho muestrales y censales, uno censal y tres no recogieron evaluaciones en este nivel. En el nivel secundario, cuatro países recogieron sólo evaluaciones censales, seis muestrales y censales, seis muestrales y sólo dos (Venezuela y Bolivia) no realizaron ninguna evaluación (Figura 21).

En los sistemas con evaluaciones censales de altas implicancias³⁶, distintas prácticas y consecuencias no previstas comenzaron a ser motivo de polémica y análisis. En México, por ejemplo, cuando se vincularon los resultados de las pruebas ENLACE a los aumentos salariales de los docentes en 2008, no se diseñó una política de monitoreo de sus consecuencias sobre la práctica docente en el aula; esto es, que la vinculación podía derivar en estrechamiento curricular o en enseñanza dirigida a la prueba. Eso implicó incentivos hacia prácticas poco éticas como la venta de exámenes o la sobrepresión de cierto tipo de estudiantes (Martínez, 2015). En Brasil, tras el surgimiento de la Prova Brasil (2005), con carácter censal para educación pública y muestral para la privada, se creó en 2007 el Índice de Desarrollo de la Educación Básica (IDEB), que marca una pauta importante en el uso de las pruebas de aprendizaje hacia la creación de indicadores sintéticos de calidad de la educación y asignación de incentivos dentro del sistema. El IDEB combina resultados de aprendizaje con tasas de aprobación, proporcionando una medida integral de la calidad educativa más allá del logro académico. La creación del IDEB introdujo un enfoque de rendición de cuentas poco común en América Latina en su época al establecer metas educativas a medio y largo plazo a distintos niveles. Ello promueve el monitoreo del progreso y de la rendición de cuentas (Ferrer y Fiszbein, 2015). Sobre la base de este primer índice, distintos estados brasileños diseñaron sistemas propios para monitorear la calidad por escuela y asignar incentivos salariales docentes corrigiendo algunas prácticas no previstas (Box 3).

36 Como decíamos anteriormente, tomamos la definición de Herrero, et al. (2022), según la cual las pruebas de altas implicancias son aquellas con consecuencias vinculantes asociadas a sus resultados para distintos actores del sistema (estudiantes, escuelas o maestros). Estas pueden ser: pruebas con fines de certificación de un curso o nivel educativo (primario o secundario) o de selección o admisión para los niveles superiores; o evaluaciones con fines de monitoreo, donde los resultados se utilizan para la rendición de cuentas por parte de las instituciones educativas o existen incentivos asociados a los resultados.

Box 3:**Equilibrar las consecuencias no previstas de las evaluaciones de aprendizajes**

Las pruebas de aprendizaje de altas implicancias han mostrado ser de gran valor en varios países del mundo cuando se combinan con buenos maestros y una visión estratégica (Banco Mundial, 2018). Al mismo tiempo, los exámenes tienen potenciales efectos negativos que han sido debatidos y estudiados, incluyendo los siguientes.

- ◆ **Reducción del plan de estudios y prácticas poco éticas.** Los exámenes de altas implicancias pueden estrechar el alcance de la educación al incentivar prácticas como enseñar para el examen, excluir aprendizajes en áreas no evaluadas o limitar la innovación en el aula, ya que los maestros podrían carecer de incentivos a probar nuevos métodos (Blazer, 2011; Ascorra et al., 2019). Las políticas fuertes de rendición de cuentas también pueden incentivar prácticas poco éticas para obtener mejores resultados en los exámenes. La literatura recoge ejemplos de exclusión de las evaluaciones a estudiantes con bajo rendimiento o con necesidades educativas especiales en Chile y de acciones fraudulentas como venta de exámenes en México (Cáceres, Muñoz y Valenzuela, 2021).
- ◆ **Distorsiones en las percepciones de calidad educativa de profesores y escuelas.** Desde muy temprano, los sistemas de evaluación educativa con exámenes de altas implicancias en la región, como el de Chile, han adoptado prácticas de publicación de resultados que permiten comparaciones con escuelas de similares características socioeconómicas. Esta información ha sido en ocasiones descontextualizada por algunos actores del sistema (por ejemplo, por la prensa o autoridades locales) y ha sido motivo de

un amplio debate debido a la estigmatización percibida de maestros y escuelas de contextos vulnerables (MINEDUCA, 2014). En Brasil, aunque el INEP evita clasificar explícitamente a las escuelas según su rendimiento promedio, los medios de comunicación y otros actores (incluidos algunos secretarios de estado) aún utilizan los resultados para construir sus propias tablas de clasificación. Aunque estas prácticas pueden tener la intención de ejercer presión interna y externa para que las escuelas mejoren, corren el riesgo de distorsionar las percepciones de la calidad educativa porque por lo general ignoran el contexto en el que opera la escuela o el progreso logrado con el tiempo (OCDE, 2021). La publicación y uso de los resultados para generar debate, discusión y presión dentro del sistema para mejorar los aprendizajes de las y los estudiantes es uno de los principales beneficios de las pruebas. Lo importante es mitigar algunos de estos riesgos ofreciendo guías sobre formas constructivas de comparar sistemas y escuelas, evitando así algunas de las consecuencias no deseadas y potencialmente negativas de tener dichos datos. La evidencia muestra que en Chile la rendición de cuentas ha tenido impactos positivos: las escuelas con bajo rendimiento han implementado políticas para mejorar a corto plazo y las crónicamente subrendidoras han invertido en apoyo técnico y formación docente para evitar sanciones (Elacqua et al., 2016; Elacqua, 2019).

- ◆ **Sobreevaluación y poca coordinación entre autoridades territoriales.** Países federales con una sólida cultura de evaluación estudiantil y un sistema nacional de evaluación establecido, como Brasil, pueden presentar problemas de ineficiencias en la coordinación de las evaluaciones a nivel local. En Brasil, alrededor de veintitrés de las veintisiete unidades federales (estados y el Distrito Federal)

han administrado sus propias evaluaciones estandarizadas en años recientes, y los municipios también pueden realizar evaluaciones para sus redes escolares. Esta disposición es necesaria en ciertos aspectos, ya que las evaluaciones locales en Brasil suelen ofrecer datos oportunos que pueden utilizarse con fines más formativos que los del SAEB. Sin embargo, la situación también presenta desafíos en términos de coordinación de políticas, ya que las y los estudiantes en varias partes del país pueden verse sometidos a múltiples evaluaciones externas que miden las mismas materias durante el mismo año escolar (OCDE, 2021).

Brasil y Chile han tomado medidas y hecho reformas específicas para atenuar algunos de los impactos negativos de las pruebas. Por ejemplo, en Brasil los estados de São Paulo y Espírito Santo buscaron fomentar la equidad en el diseño de los índices de calidad educativa al favorecer el rendimiento de estudiantes menos favorecidos. En São Paulo se crearon cuatro niveles de rendimiento del estudiante, desde básico hasta avanzado, y se calculó la proporción de estudiantes en cada nivel para enmarcar el índice de rendimiento en términos de la discrepancia entre la posición media y la ideal de cada estudiante. El objetivo era evitar que los maestros se concentraran en las y los estudiantes más propensos a mostrar mejoras en los resultados de las pruebas, dando así menos atención a aquellos en los extremos más altos y bajos del logro. En Espírito Santo se creó un nivel adicional de rendimiento del estudiante titulado “los excluidos”; este nivel comprende a todos las y los estudiantes que abandonaban la escuela o que no se presentaban para las evaluaciones de fin de año y a los que se les asignaba un peso de cero. Aquí se buscaba promover los esfuerzos para evitar la deserción y desalentar el ocultamiento de las y los estudiantes con rendimiento más bajo antes de las pruebas

(Brooke, 2016). Si bien este tipo de prácticas pueden ser útiles para contrarrestar posibles consecuencias de la vinculación de los salarios docentes al desempeño en la prueba, aún persisten varias discusiones en torno a sus consecuencias en el estrechamiento del currículo y a la limitación de prácticas en el aula hacia la preparación para los exámenes. En Chile, la inversión en el mejoramiento del sistema de monitoreo ha mostrado ser efectivo para contrarrestar la exclusión de estudiantes desaventajados en las evaluaciones (Pérez y Maldonado, 2015).

Aunque estos esfuerzos son importantes, los encargados de formular políticas deben equilibrar el peso de los exámenes en el sistema de evaluación de un país. En este sentido es importante tomar medidas para promover la complementariedad en el propósito y diseño de evaluaciones estandarizadas en los niveles nacional y subnacional para hacer un uso eficiente de los recursos, no sobrecargar a las y los estudiantes, no enviar información descontextualizada sobre los resultados de los exámenes y desincentivar prácticas poco éticas o que afecten a grupos poblacionales especiales.

Consolidación de los sistemas nacionales de evaluación (2011-2019)

La segunda década del siglo XXI se caracterizó por sistemas de evaluación más consolidados y por la aparición de nuevas entidades evaluativas que aplicaban evaluaciones censales para ambos niveles. Se adoptó una mirada crítica y analítica sobre la conceptualización de la calidad educativa, el uso de los resultados de las pruebas de aprendizaje y la aplicación de cuestionarios de clima escolar, acciones y actitudes y socioemocionales.

A nivel institucional, la década se caracterizó por el surgimiento de varias agencias independientes de evaluación de la calidad con repercusiones específicas dentro de los sistemas. Compartiendo la misión de evaluar aprendizajes, las agencias independientes de la región difieren en varias dimensiones incluyendo sus modelos de financiación, autonomía, los grados a evaluar, sus modelos de gobernanza, el alcance frente a los procesos de entrega y divulgación de resultados y su influencia en las políticas educativas. En 2011 Chile descentralizó la Agencia de la Calidad Educativa-ACE; en 2012 Ecuador creó el Instituto Nacional de Evaluación Educativa-INEVAL; en 2017 Paraguay creó el Instituto Nacional de Evaluación Educativa-INEE y Bolivia creó el Observatorio Plurinacional de la Calidad Educativa-OPCE, una institución autónoma encargada de la evaluación educativa.

Cada una de las agencias tiene funciones específicas y desarrolla distintos tipos de pruebas. En Colombia el ICFES y en Brasil el INEP están a cargo de todas las evaluaciones, incluyendo SABER 11 y SABER Pro en Colombia y SAEB y ENEM en Brasil. En Chile la evaluación para la educación superior (anteriormente PSU, ahora Prueba de Acceso a la Educación Superior, PAES) está a cargo del DEMRE mientras que la evaluación docente sigue siendo responsabilidad del CPEIP. La

Agencia de Calidad de la Educación en Chile y el INEE de Uruguay se enfocan en las pruebas estandarizadas mientras que la ANEP en Uruguay se encarga de las evaluaciones formativas y pruebas internacionales.

Los modelos de financiación de los institutos de evaluación independientes son variados. INEVAL (Ecuador), la Agencia de Calidad de Educación (Chile) e INEE (Uruguay) dependen de una asignación del presupuesto general de cada uno de los países. El ICFES (Colombia) fue constituido como una empresa social del estado que debe generar sus propios recursos para sostenerse, lo cual hace a partir de los ingresos generados por Saber 11, Saber TyT y Saber Pro; otras entidades (como el Ministerio de Educación) pueden contratar al ICFES para el diseño y aplicación de otras pruebas. En concreto, gracias a las reformas introducidas en 2009 el ICFES asumió responsabilidades más formalizadas para ofrecer servicios de evaluación, generar ingresos y utilizar los superávits obtenidos para reinvertir en el desarrollo técnico y programático (Ferrer y Fiszbein, 2015). El INEP (Brasil) tiene un modelo mixto: percibe ingresos directamente de las personas que toman el ENEM y recibe una partida presupuestaria para el censo escolar y las pruebas SAEB. En todos los casos hay discrecionalidad por parte de las autoridades frente a la decisión de aplicar las pruebas y de asignar los recursos necesarios para hacerlo a excepción de aquellas que son pagadas por los ciudadanos (Saber 11, Saber TyT, Saber Pro y ENEM). El marco institucional de cada país determina la estructura del instituto de evaluación independiente, en caso de considerarlo necesario, pues no es la única alternativa.

Más allá de los distintos modelos, pocos institutos de evaluación en la región han logrado blindarse frente a cambios en decisiones de política educativa. Por ejemplo, el Observatorio Plurinacional de la Calidad Educativa-OPCE de Bolivia ha enfrentado dificultades técnicas y políticas para poder llevar a cabo acciones de medición de los aprendizajes luego de que cambiara el contexto político del país³⁷. En México, luego de la reforma de educación de 2013, y la subsecuente implementación en 2018, desapareció el INEE, que había sido el organismo técnico que se ocupaba de las evaluaciones de logro educativo, lo que presentó un retroceso en la realización de las pruebas³⁸.

A pesar de los cambios institucionales, el nivel de aplicación de las pruebas se mantuvo con respecto a la década anterior. En esta década aumentó el número de países con pruebas en primaria debido a la implementación de pruebas diagnósticas muestrales en República Dominicana y disminuyó el número de pruebas en secundaria debido a la ausencia de aplicación de pruebas en Panamá (Figura 19). Por tipo de evaluación, Nicaragua y Uruguay comenzaron a aplicar pruebas censales en secundaria, Panamá en primaria y Argentina y Paraguay en ambos niveles. De esta forma, 8 de los 18 países contaban con pruebas censales en primaria y trece en secundaria (Figura 21).

Por otro lado, hubo una amplia discusión sobre la conceptualización de la calidad educativa, los usos de las pruebas y el rumbo de la política pública. En esta década surgieron discusiones respecto a la calidad educativa como un proceso de obtención de resultados en pruebas a gran escala que sólo daban cuenta de resultados en procesos cognitivos (Martínez-Iñiguez, Tobón y López-Ramírez, 2020; Vázquez Olivera, 2015). Bajo este contexto, varios países dieron importancia a medir otro tipo de habilidades además del monitoreo de áreas básicas y se crearon y utilizaron en mayor medida índices de calidad educativa para medir de forma sintética y precisa la calidad del servicio que proporcionan las escuelas y, al mismo tiempo, guiar prácticas de escuelas, distritos, directivos y docentes (Elacqua, Martínez y Westh, 2019). Dado que en la mayoría de los países de la región existían, dentro de los estándares de aprendizaje, dimensiones y habilidades relacionadas con el desarrollo socioemocional, ya sea de forma transversal o en una asignatura o de tiempo específico en el currículo, varios países comenzaron a recoger en formularios anexos a sus evaluaciones a gran escala preguntas para medir distintas habilidades socioemocionales de los y las estudiantes tanto para su mejora o como insumo para el análisis de los resultados de desempeño (Arias, Hincapié y Paredes, 2020).

37 De acuerdo con Zea (2020), con la implementación de la Ley Educativa Avelino Siñani-Elizardo Pérez N° 070 (ASEP) la política educativa en Bolivia se caracterizó por un discurso contracorriente de las tendencias globales, la reducción de la inscripción en escuelas privadas y el alejamiento de las evaluaciones a gran escala internacionales.

38 De acuerdo con Flamand, Arriaga y Santizo (2020), después de la reforma de 2013 el INEE se convirtió en la autoridad principal en la evaluación de política educativa, traslapando las funciones de la Secretaría de Educación Pública-SEP. En su nuevo rol, el INEE era el encargado no sólo de la evaluación sino de proponer políticas para fomentar la calidad educativa y docente. Los docentes se mostraron muy disconformes por la reforma y por el peso que las evaluaciones a gran escala tenían sobre su calificación de rendimiento; sumado a errores de comunicación del INEE y a disconformidades también de la SEP, esto llevó a una revuelta política importante que terminó con la disolución del instituto.

Box 4:**Medición de habilidades socioemocionales en América Latina**

En la segunda década del siglo comenzó en la región una amplia discusión en torno a la conceptualización y operativización de la medición de habilidades socioemocionales.

Hasta 2019 sólo Chile, Colombia, México y Uruguay incluían dentro de las evaluaciones a gran escala formularios adicionales para su medición como autorreporte. Aunque estos fueron avances importantes, ha existido cierta falta de claridad conceptual para nombrar y operacionalizar las habilidades socioemocionales; los países han adoptado distintos marcos y conceptos como el

aprendizaje socioemocional, las habilidades no cognitivas, las habilidades del siglo XXI o las habilidades blandas, entre otros (Arias, Hincapié y Paredes, 2020; INEE, 2019).

A raíz de la pandemia del COVID-19, Colombia implementó cambios en el marco de la medición de habilidades socioemocionales para ofrecer un marco general de su proceso de formación y poder evaluar los resultados de los esfuerzos de las escuelas por alcanzar la meta de formar a sus estudiantes en habilidades socioemocionales (ICFES, 2021). Ecuador y Perú comenzaron a medir sistemáticamente este tipo de habilidades también bajo formularios aplicados a las y los estudiantes.

	Chile	Colombia	México	Uruguay	Ecuador	Perú
Primera medición	2013	2002*	2017	2017	2021	2021
Última medición	2023	2022	2019	2023	2023	2022
Marco / reforma		2021		2017		
Contenido de formulario socioemocional						
Manejo de situaciones adversas			✓		✓	
Confianza en sí mismo					✓	
Auto percepción	✓				✓	
Autovaloración académica y motivación escolar	✓			✓		
Autoeficacia				✓		✓
Autoconfianza		✓				
Satisfacción con la vida						
Autorregulación conductual				✓		✓
Autonomía y relación						✓
Empatía			✓	✓		✓
Responsabilidad						✓
Toma de decisiones			✓			✓
Reconocimiento de emociones		✓				
Expresión de emociones		✓				
Trabajo en equipo		✓				
Sentido de agencia			✓			
Perseverancia	✓		✓	✓		
Regulación emocional			✓	✓		
Habilidades de relacionamiento						
Habilidades de redes sociales			✓			
Mentalidad de crecimiento			✓			
Valoración de la tarea				✓		
Conductas de riesgo (inter o externalizante)				✓		
Grados	4° y 11° (anual), 6° y 8° (cada 2 años)	3°, 5°, 9°, desde 7°	6°, 9°, 12°	6°, 9°	4° primaria, 3° secundaria	6°, 2° secundaria
Tipo de prueba	Censal	Censal hasta 2018, muestral desde 2022.	Censal / muestral	Muestral	Muestral / censal	Censal / muestral

Se hizo más común la medición de habilidades socioemocionales en la región.

En Colombia la implementación de formularios para obtener información de acciones y actitudes, clima escolar y socioemocionales de las y los estudiantes comenzó en el marco de la prueba de Competencias Ciudadanas, con aplicaciones desde 2002³⁹ para las y los estudiantes de quinto grado de primaria y noveno de secundaria. El objetivo era medir la autorregulación emocional de las y los estudiantes, su automotivación, autoeficiencia escolar y trabajo en equipo. La prueba de aprendizajes PLANEA en México, además de evaluar las áreas de lenguaje y matemáticas, incluyó un cuestionario para medir las habilidades socioemocionales de las y los estudiantes desde 2017. En sexto y noveno grado, el cuestionario medía las habilidades de empatía, sentido de agencia y regulación emocional en línea con las habilidades socioemocionales desarrolladas en el Plan y Programas de Estudios en Educación Básica (primero a noveno grado). En grado doce, el cuestionario medía empatía, perseverancia, manejo de estrés y toma de decisiones (Arias, Hincapié y Paredes, 2020). En Uruguay, la evaluación de logros educativos Aristas, además de medir el desempeño en lectura y matemáticas, incluye un cuestionario para medir las habilidades socioemocionales de los y las estudiantes desde 2018.

Varios países crearon y utilizaron índices sintéticos de la calidad basados en los resultados de las pruebas y otras características de los sistemas.

En Colombia a fines de 2014 se diseñó el Índice Sintético de la Calidad Educativa (ISCE) para poner a disposición de la comunidad educativa un indicador de calidad de escuelas que fuese fácil de interpretar y que sirviese como insumo para la reflexión y definición de metas específicas a cada establecimiento según el nivel educativo ofrecido (primaria, secundaria y media). Su elaboración tuvo como principal referencia el IDEB de Brasil (2007) y tomó en cuenta los niveles de aprendizaje de las y los estudiantes en la escuela, la tasa de aprobación y reten-

ción, mejoras en el aprendizaje e indicadores de ambiente escolar recogidos en los formularios de contexto de las pruebas. Desde 2013, Chile mide los indicadores de desarrollo personal y social (IDP) que forman parte de las distintas dimensiones de “otros indicadores de calidad educativa”: autoestima académica y motivación escolar, participación y formación ciudadana, clima de convivencia escolar. En Chile desde 1995 se han creado rankings de las escuelas de acuerdo con criterios de evaluación que han variado en el tiempo. En 2011 se adoptó una nueva metodología para el ordenamiento de escuelas según su calidad, descrita en la Ley de Aseguramiento de la Calidad. El objetivo principal del índice es responsabilizar a las escuelas y a sus directivos por el cumplimiento de los estándares de aprendizaje y la calidad del servicio que ofrecen e identificar necesidades de apoyo focalizado. Para esto se pondera el nivel de aprendizaje de las y los estudiantes de las escuelas (pruebas), la tasa de aprobación y retención, mejoras en el aprendizaje, medidas de equidad de las escuelas e indicadores de ambiente escolar. Si bien cada uno de los índices en Brasil, Colombia y Chile tiene objetivos de ordenamiento diferenciados, entre 50 y 80 por ciento del peso de los indicadores de calidad de escuelas corresponde al nivel y mejoras de aprendizaje medidas por estas evaluaciones (Elacqua, Martínez y Westh, 2019).

En esta década también se hizo más común la publicación de los microdatos, formularios e informes técnicos de las pruebas dentro de las plataformas de los ministerios o de las agencias de calidad. Esta práctica ha buscado estimular la transparencia y ha potenciado el uso de los resultados con fines de investigación y políticas públicas. Los microdatos son publicados en 10 de los 18 países (Argentina, Brasil, Colombia, Ecuador, Guatemala, México⁴⁰, Paraguay y Uruguay) o pueden ser requeridos dentro de las plataformas de las instituciones (Perú y Chile).

39 Hasta 2006, exceptuando 2004, y luego una única aplicación en 2012.

40 Antes de la disolución del INEE.

Box 5: las evaluaciones nacionales de aprendizaje en el Caribe

Además de las evaluaciones realizadas en el marco del Caribbean Examinations Council (CXC), varios países del Caribe llevan a cabo

sus propias evaluaciones nacionales para evaluar el aprendizaje de las y los estudiantes en diversas áreas y niveles educativos. A continuación presentamos una breve síntesis sobre las evaluaciones nacionales de los ocho países del Caribe miembros del BID.

País	Evaluación	Nivel educativo	Grado	Descripción
Bahamas	Grade level assessment test (GLAT)	Primaria	3° y 6°	Evalúa matemáticas y artes del lenguaje. Las y los estudiantes de sexto grado también reciben pruebas en ciencias y estudios sociales. El GLAT se administra a todas las escuelas primarias gubernamentales y a la mayoría de las escuelas primarias independientes.
	Bahamas Junior Certificate (BJC)	Secundaria baja	9°	Evalúa el dominio de materias básicas y es requerido para avanzar al siguiente ciclo (secundaria alta). Conduce a la obtención del Certificado Junior de las Bahamas (BJC), está diseñada para medir el dominio del plan de estudios en materias como arte, manualidades, idioma inglés, literatura inglesa, ciencias generales, ciencias de la salud, economía doméstica, matemáticas, educación religiosa, estudios sociales y dibujo técnico.
	Bahamas General Certificate of Secondary Education (BGCSE)	Secundaria alta	12° (último año)	Examen de salida necesario para la universidad o estudios profesionales. Se basa en los modelos del Certificado General de Educación Secundaria (GCSE) del Reino Unido y del Certificado General de Educación Secundaria Internacional (IGCSE), pero se adapta para satisfacer las necesidades específicas de las Bahamas. Se ofrecen actualmente 25 materias.
Barbados	Barbados Secondary Schools' Entrance Examination (BSSEE)	Primaria	6°	Se utiliza para asignar estudiantes a escuelas secundarias y evalúa el nivel de competencia de las y los estudiantes en una variedad de áreas de estudio.
Belice	ATLIB	Secundaria alta	12° (último año)	Examen al final de la secundaria, necesario para acceso a instituciones de nivel terciario.
	Belize Junior Achievement Test (BJAT)	Primaria	6°	Evalúa habilidades en matemáticas, inglés, ciencias sociales y naturales. La evaluación fue suspendida en 2019/2020 debido a la pandemia; desde entonces no se ha vuelto a implementar.
	Primary School Examinations (PSE)	Secundaria baja	8°	Evalúa matemáticas, inglés, ciencias sociales y naturales. La evaluación fue suspendida en 2019/2020 debido a la pandemia; desde entonces no se ha vuelto a implementar.
Jamaica	Grade One Individual Learning Profile (GOILP)	Primaria	1°	El GOILP se administra individualmente antes de que las y los estudiantes ingresen al primer grado. La evaluación proporciona información inicial para cada estudiante, incluidas las habilidades y nivel de competencia de las y los estudiantes en seis subconjuntos: conocimientos generales; conceptos numéricos; lenguaje oral, lectura y escritura; dibujo; hábitos de trabajo; y comportamiento en el aula.
	Grade 3 Diagnostic Tests (G3DT)	Primaria	3°	Se basa en el plan de estudios y se administra a estudiantes que han completado el tercer grado. Se aplica en lenguaje y matemáticas al final del año escolar (junio). Su objetivo principal es identificar áreas de debilidad en el aprendizaje de las y los estudiantes.
	Grade Four Literacy Test (GFLT) Grade Six Achievement Test (GSAT)	Primaria	4°	El componente de alfabetización incluye la evaluación del reconocimiento de palabras, comprensión de lectura y escritura. Para la parte de numeración, las y los estudiantes son evaluados en operaciones y representación numérica, medidas y geometría y álgebra y estadística. Se asigna uno de tres niveles: dominio, casi dominio o no dominio. Las y los estudiantes que no logran el dominio tanto en alfabetización como en numeración no pueden presentar el PEP y deben volver a realizar los exámenes.
	Primary Exit Profile (PEP)	Primaria	4°, 5° y 6°	Es una serie de evaluaciones que se realizan a lo largo de cuarto, quinto y sexto grado, comenzando con el GFLT y GFNT en el cuarto grado. Su objetivo es medir el nivel de preparación para el séptimo grado y ubicar a las y los estudiantes en escuelas secundarias. Evalúa habilidades en artes del lenguaje y matemáticas (en los grados cuarto, quinto y sexto) y estudios sociales y ciencias (sólo en quinto grado). La serie de evaluaciones consta de tres componentes: tareas de rendimiento (en los tres niveles), pruebas de habilidad y pruebas basadas en el plan de estudios (sólo para aquellos en el sexto grado).

País	Evaluación	Nivel educativo	Grado	Descripción
Guyana	National Grade Two Assessment (NGTA)	Primaria	2°	Evaluación diagnóstica basada en el plan de estudios de lenguaje y matemáticas de primero y segundo grado. Contribuye 5 por ciento a los resultados de NGSA (National Grade Six Assessment).
	National Grade Four Assessment (NGFA)	Primaria	4°	Evaluación diagnóstica fundamentada en el plan de estudios de primero a cuarto grado; incluye alfabetización y numeración. Contribuye 10 por ciento a los resultados de NGSA.
	National Grade Six Assessment (NGSA)	Primaria	6°	Utilizada para la asignación de estudiantes a escuelas secundarias. Evalúa matemáticas, lenguaje, ciencias y estudios sociales. Las y los estudiantes deben haber completado previamente NGTA y NGFA.
	National Grade Nine Assessment (NGNA)	Secundaria baja	9°	Evaluación diagnóstica basada en el plan de estudios de séptimo a noveno grado.
Trinidad y Tobago	Secondary Entrance Assessment (SEA)	Primaria	5°	Se realiza al finalizar quinto grado, aproximadamente a los 11 años; es utilizada para la asignación de vacantes en escuelas secundarias.
	National Certificate of Secondary Education (NCSE)	Secundaria baja	8°	Necesario para el ingreso a la educación secundaria alta. Se realiza al final del tercer ciclo en escuelas públicas y privadas. Abarca ocho materias (matemáticas, inglés, español, ciencias, estudios sociales, artes, educación física y estudios técnicos).
Surinam	GLO 6	Primaria	6°	Evalúa el nivel de conocimientos y habilidades de las y los estudiantes al final de la educación primaria. Se evalúa matemáticas, lengua materna, ciencias sociales y ciencias naturales. Los que obtienen notas aprobatorias tienen la posibilidad de matricularse en el primer ciclo académico de educación secundaria; los restantes pueden optar por la educación secundaria técnica o profesional.
Haití	Evaluation Nationale d'apprentissages (ENA)	Primaria	4° (2016-2017) y 6° (2022)	Tras un pre-piloto en 2015 en 60 escuelas, se llevaron a cabo las primeras dos rondas de evaluaciones nacionales en cuarto grado en 2016 y 2017, sustituyendo a las anteriores pruebas de sexto grado. Buscan proporcionar información sobre el nivel de competencia de las y los estudiantes en matemáticas, creole y francés antes de finalizar la educación primaria. La aplicación de 2016 fue muestral, con 526 escuelas a nivel nacional; la de 2017 fue censal para escuelas públicas (1.771) y muestral para escuelas privadas (204). Para la evaluación de 2022 se optó por realizarla en sexto grado, manteniendo las mismas asignaturas y un diseño muestral a nivel nacional.

Disrupción y estado actual de las pruebas (2020-2023)

La disrupción que generó el COVID-19 afectó la continuidad de las pruebas en la mayoría de los países. La paulatina recuperación de los relevamientos permite reflexionar sobre la importancia de sus objetivos y su institucionalidad, sobre la consolidación del sistema de evaluación y financiamiento y en la estabilidad de su recolección y uso. La disrupción ha acelerado la discusión sobre las tendencias y retos hacia el futuro de las evaluaciones en los sistemas educativos.

La disrupción del COVID-19 impactó sobre la realización de las pruebas. Sólo tres países de la región continuaron con las pruebas en 2020. Durante la pandemia el proceso educativo de los niños, niñas y jóvenes en la región se vio fuertemente afectado, recordando que en la región se dieron los cierres más prolongados de escuelas a nivel mundial (Herrero et al., 2022). Esta disrupción también afectó la realización de las evaluaciones a gran escala nacionales, que eran recogidas de forma presencial en papel y lápiz en casi todos los países. La excepción es Uruguay, que desde 2013 había acumulado experiencia en la aplicación de pruebas digitales como la prueba Nacional de Aprendizajes 2013 y los pilotos digitales de Aristas 2017 y 2019). En 2020 sólo Colombia y El Salvador recogieron pruebas censales a ni-

vel de secundaria y Uruguay recogió la prueba Aristas (muestral) a nivel primaria. De los 18 países analizados, sólo tres recogieron algún tipo de evaluación nacional a gran escala.

La continuidad de las pruebas se vio afectada durante la pandemia, pero se evidencia una rápida recuperación. A nivel mundial, en el año escolar 2021/2022, al menos el 70 % de los países continuaron con sus programas de evaluaciones nacionales a gran escala (OCDE, UNESCO, UNICEF y Banco Mundial, 2022). En América Latina, el porcentaje de países que reanudaron las evaluaciones entre 2021 y 2023 superó el promedio global, alcanzando el 78 %. En el nivel primario, 11 de los 18 países analizados aplicaron al menos una prueba entre 2021 y 2023. Sin embargo, en cuatro de los siete países se dejaron de aplicar pruebas muchos antes de la pandemia: en Bolivia y Venezuela, la última evaluación a gran escala en primaria se realizó hace más de 20 años, mientras que en Nicaragua y El Salvador hace más de 8 años. En Uruguay las pruebas Aristas de primaria aún se encuentran dentro del ciclo presupuestado de recolección considerando que su última aplicación fue en 2020 y la frecuencia establecida es de cada tres años. En secundaria, 14 de los 18 países analizados llevaron a cabo evaluaciones a gran escala después del 2020. En los restantes cuatro países – Bolivia, Panamá, Nicaragua y Venezuela - estas evaluaciones no se realizan desde hace al menos 8 años (Tabla 2).

Tras la pandemia hubo cambios en la cobertura y niveles recolectados. Dentro de los países que aplicaron pruebas a nivel de primaria, tras la pandemia pasaron de ser censales a muestrales en Colombia y Paraguay, mientras que en Costa Rica se implementaron de nuevo pruebas censales para primaria. Como discutiremos más adelante, hubo factores presupuestarios y políticos para estos cambios. Argentina pasó de recoger pruebas para tercero y sexto grado de primaria a sólo hacerlo en sexto, y Chile de cuarto y sexto a sólo sexto. En secundaria, la prueba de Perú pasó a ser muestral; Ecuador pasó de recoger la prueba en décimo de secundaria y tercero en media a sólo hacerlo en décimo; y los sistemas de El Salvador, Guatemala y México también disminuyeron el número de grados evaluados, dejando sólo evaluaciones en el último grado del ciclo. Por el contrario, Colombia comenzó a levantar por primera vez pruebas en séptimo de secundaria.

Tabla 2: aplicación de evaluación(es) por año, 1990 – 2023 18 países

Primaria CINE 1

	ARG	BOL	BRA	CHL	COL	CRI	DOM	ECU	GTM	HND	MEX	NIC	PAN	PRY	PER	SLV	URY	VEN
90			✓															
91																		
92				✓														
93			✓	✓														
94				✓														
95	✓		✓	✓														
96	✓			✓				✓						✓	✓		✓	
97	✓	✓	✓					✓		✓				✓				
98	✓							✓	✓	✓	✓				✓		✓	✓
99	✓		✓	✓		✓			✓	✓	✓			✓			✓	
00	✓					✓		✓	✓	✓	✓			✓				
01			✓			✓				✓	✓				✓	✓	✓	
02				✓		✓				✓	✓	✓					✓	
03	✓		✓		✓	✓				✓	✓							
04						✓			✓	✓				✓	✓			
05	✓		✓	✓	✓	✓				✓	✓		✓			✓	✓	
06				✓		✓				✓	✓	✓						
07	✓		✓	✓				✓	✓	✓	✓				✓			
08				✓		✓		✓	✓	✓					✓	✓		
09			✓	✓	✓			✓	✓	✓	✓	✓			✓		✓	
10	✓			✓				✓	✓	✓	✓	✓		✓	✓			
11			✓	✓				✓		✓	✓				✓			
12				✓	✓	✓		✓		✓	✓				✓	✓		
13	✓		✓	✓	✓			✓	✓	✓	✓				✓		✓	
14			✓	✓	✓			✓	✓	✓	✓				✓			
15			✓	✓	✓			✓		✓	✓	✓		✓	✓			
16	✓		✓	✓	✓			✓		✓			✓		✓			
17	✓		✓	✓	✓		✓	✓		✓			✓				✓	
18	✓			✓			✓	✓		✓	✓		✓	✓	✓			
19			✓	✓		✓		✓	✓		✓				✓			
20																	✓	
21	✓		✓		✓	✓		✓										
22	✓			✓	✓		✓	✓	✓						✓			
23	✓		✓	✓	✓	✓	✓	✓	✓	✓				✓	✓			

Secundaria CINE 2 y 3

	ARG	BOL	BRA	CHL	COL	CRI	DOM	ECU	GTM	HND	MEX	NIC	PAN	PRY	PER	SLV	URY	VEN
90					✓													
91					✓		✓											
92				✓	✓		✓											
93	✓			✓	✓		✓											
94	✓			✓	✓		✓											
95	✓		✓	✓	✓	✓	✓											
96	✓			✓	✓	✓	✓	✓										
97	✓		✓	✓	✓	✓	✓	✓								✓		
98	✓			✓	✓	✓	✓	✓						✓	✓	✓		✓
99	✓		✓		✓	✓	✓									✓	✓	
00	✓			✓	✓	✓	✓	✓								✓		
01			✓	✓	✓	✓	✓								✓	✓		
02					✓	✓	✓									✓		
03	✓		✓	✓	✓	✓	✓									✓	✓	
04				✓	✓	✓	✓								✓	✓		
05	✓		✓	✓	✓	✓	✓				✓		✓			✓		
06				✓	✓	✓	✓		✓					✓		✓		
07	✓		✓	✓	✓	✓	✓	✓	✓	✓				✓		✓		
08				✓	✓	✓	✓	✓	✓	✓	✓					✓		
09			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓				✓		
10	✓			✓	✓	✓	✓	✓	✓	✓	✓	✓		✓		✓		
11			✓	✓	✓	✓	✓	✓	✓	✓	✓					✓		
12				✓	✓	✓	✓	✓	✓	✓	✓					✓		
13	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓					✓		
14				✓	✓	✓	✓	✓	✓	✓	✓					✓		
15			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		
16	✓			✓	✓	✓	✓	✓	✓	✓					✓	✓		
17	✓		✓	✓	✓	✓		✓	✓	✓	✓					✓		
18				✓	✓	✓		✓	✓	✓				✓	✓	✓	✓	
19	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓				✓	✓		
20					✓											✓		
21			✓		✓	✓		✓	✓							✓		
22	✓			✓	✓		✓	✓	✓		✓				✓	✓	✓	
23			✓	✓	✓	✓	✓	✓	✓	✓				✓	✓	✓		

Fuente: elaboración propia sobre la base de la revisión de la información de cada sistema.

Objetivo o implicancia de las pruebas

Como hemos mencionado, una forma de clasificar las pruebas de aprendizaje es de acuerdo con las implicancias de sus resultados. Hay pruebas de bajas implicancias que generan información para fines formativos de diverso tipo; pueden ser censales o muestrales y no tienen consecuencias directas para los actores involucrados. Por otro lado, las pruebas de altas implicancias usan los resultados para diversos fines. Por ejemplo, pueden usarse para definir incentivos para los profesores y escuelas (Chile, Brasil, Perú y antes México), para la promoción educativa de las y los estudiantes (Costa Rica, El Salvador, República Dominicana y Colombia para el acceso a educación superior) o informar a la población sobre el rendimiento de las y los estudiantes en los centros educativos (Brasil, Chile, Colombia). Los sistemas de altas implicancias a menudo utilizan marcos de evaluación censales y usualmente forman parte de un sistema de rendición de cuentas en educación (Cueto, 2007).

Antes de la pandemia, 16 de los 18 países recogieron pruebas en el nivel secundario de educación. De estos, siete recogieron pruebas censales de altas implicancias para las escuelas o docentes (SIMCE en Chile, Prova en Brasil y pruebas Censales en Perú) o servían como indicador de certificación o aprobación para las y los estudiantes de educación media (PAES en El Salvador, Bachiller en Costa Rica, Pruebas Nacionales en República Dominicana) o como prueba de ingreso a la educación superior (Saber 11 en Colombia). Durante la pandemia, las pruebas de ingreso a la educación superior de Colombia y de certificación de El Salvador se realizaron, ajustando en ambos casos los calendarios y pasando la prueba de papel y lápiz a digital en el caso de El Salvador⁴¹. En Costa Rica, en contraste, la prueba Bachillerato había tenido vigencia desde 1988 hasta 2019, cuando el Consejo Superior de Educación (CSE) aprobó las Pruebas Nacionales para el Fortalecimiento de Aprendizajes para la Renovación de Oportunidades (FARO),

con el fin de determinar el nivel de logro esperado por las y los estudiantes (primaria, técnica y media) y como requisito de promoción. Con motivo de la pandemia y por factores políticos, el comienzo de la prueba se postergó hasta 2021, se canceló de forma temporal en 2022 (MEP, 2022), para luego retomar en 2023.

En general, una de las características principales de las pruebas de altas implicancias es su carácter censal y su estabilidad en el tiempo. Las pruebas de altas implicancias para las y los estudiantes usualmente se administran de acuerdo con un calendario riguroso y sus resultados son entregados a escuelas y estudiantes en fechas previamente anunciadas en el calendario académico; algunas forman parte del avance del proceso académico de las y los estudiantes, por lo cual su continuidad es altamente importante (Ferrer y Fiszbein, 2015). En El Salvador las pruebas Avanzo (2021-2023) reemplazaron a las pruebas PAES, que habían sido realizadas anualmente desde 1997. Igualmente, las pruebas Saber 11 en Colombia han sido realizadas anualmente de forma continua desde 1968. A nivel de primaria, las pruebas SIMCE en Chile no tienen altas implicancias para las y los estudiantes pero sí las tienen para las escuelas y también se han caracterizado por una vigencia importante (con distintas temporalidades y características) desde 1992.

41 En 2020, la prueba PAES (1997-2019) fue reemplazada por la prueba Avanzo (2020-2023). Dentro de sus diferencias se encuentra la inclusión de un componente socioemocional en Avanzo y una forma de recolección digital (Campos Morán, Navarrete Gálvez y Campos Solórzano 2023).

Tabla 3: tipo de prueba e implicancias antes, durante y luego de la interrupción de la pandemia

	2011 - 2019		2020		2021 - 2023		2011 - 2019		2020		2021 - 2023	
	Altas	Bajas	Altas	Bajas	Altas	Bajas	Altas	Bajas	Altas	Bajas	Altas	Bajas
ARG		Muestrales				Muestrales		Muestrales				Censales
BOL												
BRA		Muestrales				Muestrales		Muestrales				Censales
CHL	Censales				Censales		Censales				Censales	
COL		Censales				Muestrales	Censales		Censales			Muestrales
CRI		Muestrales			Censales		Censales				Censales	
ECU		Muestrales				Muestrales		Muestrales				Muestrales
SLV		Muestrales						Muestrales	Censales		Censales	
GTM		Muestrales				Muestrales		Censales				Muestrales
HND		Muestrales				Muestrales		Muestrales				Muestrales
MEX		Muestrales						Muestrales				Muestrales
NIC		Muestrales						Muestrales				
PAN		Muestrales										
PRY		Censales				Muestrales		Censales				Muestrales
PER		Muestrales				Muestrales		Censales				Muestrales
DOM		Muestrales				Muestrales	Censales	Muestrales			Censales	Muestrales
URY		Muestrales		Muestrales				Muestrales				Muestrales
VEN												
	Primaria						Secundaria					

Censales Muestrales

Fuente: elaboración propia sobre la base de la revisión de la información de cada sistema.

Institucionalidad del sistema de evaluación

Como comentamos, el entorno institucional en el cual se llevan a cabo las evaluaciones (la dependencia o independencia de la institución responsable de la realización de las evaluaciones respecto al ministerio de educación) es un factor que afecta la continuidad de su aplicación, cobertura y uso. Esto incluye la estabilidad de los equipos técnicos, la asignación y gestión de recursos humanos y presupuestarios, la comunicación entre evaluadores y usuarios de la información y la gestión política del discurso en torno a las evaluaciones, entre otros aspectos (Galas, Gutiérrez y Hamilton, 2020; Ferrer y Fiszbein, 2015).

Una diferencia más o menos consistente es que las instituciones independientes tienen la capacidad de diseñar y llevar a cabo planes de evaluación a mediano y largo plazo y autonomía en la divulgación de sus resultados. Para los organismos no independientes puede ser un reto mantener planes de evaluación a largo plazo debido a las demandas de autoridades políticas, sobre todo cuando hay cambios de gobierno. En Costa Rica los últimos dos cambios de gobierno han impac-

tado sobre la continuidad de las pruebas estandarizadas. Entre 2019 y 2020 se eliminaron las Pruebas Nacionales Bachillerato vigentes desde 1988, que fueron sustituidas por las pruebas FARO; y tras el cambio de gobierno en 2022 las pruebas FARO fueron eliminadas y sustituidas por la Prueba Nacional Estandarizada en 2023 (Consejo Nacional de Rectores, 2023). En Venezuela y Bolivia factores políticos han afectado la consolidación de los sistemas nacionales de evaluación; Bolivia sólo ha realizado un relevamiento para primaria, en 1997, y Venezuela sólo ha realizado un relevamiento para ambos niveles en 1998.

Los sistemas con instituciones independientes presentan fortalezas como amplias trayectorias y continuidad en las pruebas debido también a una mayor autonomía financiera que promueve su sostenibilidad y mayor especialización técnica. Por otro lado, estos institutos presentan retos en la definición de los grados de autonomía y de los mecanismos de coordinación con el ministerio de educación y otras instituciones, la alineación de las pruebas con el currículo y la vinculación de estrategias de orientación y mejora. En este sentido, hay países cuyas unidades dentro de

los ministerios de educación han presentado una consistencia importante. Así es en Argentina, Perú y Guatemala, que han dejado de depender de presupuestos por acuerdos de

cooperación con organismos internacionales y cuyas evaluaciones han pasado a formar una parte importante del presupuesto educativo nacional (Ferrer y Fiszbein, 2015).

Tabla 4: presencia de evaluación por año y tipo de institución evaluadora, 2000 – 2023 18 países

Primaria CINE 1

Año	00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	
Autónomas	BRA		✓		✓		✓		✓		✓		✓		✓	✓	✓		✓		✓		✓		
	BOL																								
	COL				✓		✓			✓			✓	✓	✓	✓	✓	✓				✓	✓	✓	
	CHL			✓			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓	
	ECU	✓							✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	
	URY		✓	✓			✓				✓				✓				✓		✓				
	PRY	✓				✓					✓						✓			✓				✓	
No autónomas	ARG	✓			✓		✓		✓		✓			✓			✓	✓	✓			✓	✓	✓	
	CRI	✓	✓	✓	✓	✓	✓		✓				✓							✓		✓		✓	
	DOM																	✓	✓				✓	✓	
	HND	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓					✓
	NIC			✓				✓			✓	✓					✓								
	VEN																								
	SLV		✓							✓				✓											
	GTM	✓				✓		✓	✓	✓	✓	✓			✓	✓					✓			✓	✓
	PAN						✓											✓	✓	✓					
	MEX	✓	✓	✓	✓		✓	✓	✓		✓	✓	✓	✓	✓	✓	✓			✓	✓				
PER		✓			✓			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓			✓	✓	

Secundaria CINE 2 y 3

Autónomas	BRA		✓		✓		✓		✓		✓		✓		✓	✓	✓		✓		✓		✓		
	BOL																								
	COL	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	CHL	✓	✓		✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓
	ECU	✓							✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓
	URY				✓															✓				✓	
	PRY						✓	✓			✓						✓			✓					✓
No autónomas	ARG	✓			✓		✓		✓		✓			✓			✓	✓		✓			✓		
	CRI	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓		✓	
	DOM	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓		✓	✓	
	HND							✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓				✓	
	NIC									✓	✓					✓									
	VEN																								
	SLV	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
	GTM							✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓
	PAN						✓																		
	MEX						✓			✓	✓	✓	✓	✓	✓	✓	✓		✓		✓			✓	
PER		✓			✓											✓	✓		✓	✓			✓	✓	

✓ Evaluación Cambios institucionales

Fuente: elaboración propia sobre la base de la revisión de la información de cada sistema.

La claridad institucional y de financiamiento de las evaluaciones es un factor importante para garantizar su continuidad. En Honduras, el proyecto de cooperación internacional USAID a través del Proyecto Mejorando el Impacto al Desempeño Estudiantil de Honduras (MIDEH) aseguró la continuidad de las pruebas entre 2005 y 2017. A pesar de que la Secretaría de Educación asignó una partida presupuestaria específica para las evaluaciones estandarizadas de los aprendizajes a partir de 2018, tras la finalización del proyecto USAID en 2018 se dejó de alimentar el banco de ítems (FEREMA, 2022), retomándose las evaluaciones nacionales en el año 2023 pero sin publicación de sus resultados. Entre 2021 y 2023 no se han recolectado pruebas: en El Salvador a nivel primaria, donde la última prueba se aplicó en 2012 con fondos internacionales; en México, donde la disolución del INEE y la posterior toma de funciones de evaluación por parte de la Comisión Nacional para la Mejora Continua de la Educación (Mejoredu) ha causado retrasos en la continuidad de las pruebas; y en Panamá, donde el ministerio ha postergado el relevamiento de las pruebas CRECER más de tres años. En Paraguay, la finalización de recursos del Fondo para la Excelencia de la Educación y la Investigación (FEEI) en 2023, que venía financiado el INEE desde 2015, pone en riesgo la continuidad de las evaluaciones a gran escala.

La disposición e incentivos hacia la utilización de la información de las evaluaciones con fines de diseminación e investigación también es un factor importante que promueve la continuidad y mejora. En Brasil, el INEP creó el Comité de Productos de Información Educativa-COMPED (1998), que reúne a quince instituciones del sistema de educación para desarrollar un sistema articulado de diseminación de información educativa. El ICFES colombiano realiza desde 2010 convocatorias de investigación para la construcción de análisis sobre la calidad de la educación en el país y promueve el uso y mejoramiento de la información que genera. Todos los sistemas de evaluación con instituciones autónomas de los países analizados (excepto Bolivia) ponen a disposición libre o bajo requerimiento (Chile)

los microdatos de las pruebas de evaluación que recogen, lo que fomenta la transparencia y el uso de la información con fines investigativos. Dentro de los países con instituciones no autónomas, Argentina, Paraguay, Perú y México (en la plataforma del extinto INEE) ponen esta información a disposición del público.

Uso de tecnologías digitales

Una de las principales características de las pruebas realizadas durante la disrupción de la pandemia en Uruguay, El Salvador y Colombia es la forma en la que estos países se habían venido adaptando al uso de tecnologías digitales para las pruebas, o cómo habían adoptado políticas de acceso a dispositivos digitales. En Uruguay, el único país de la región en mantener una evaluación de implicancias bajas para primaria, factores propios de la prueba y externos influyeron sobre su recolección. Las pruebas Aristas son muestrales y en formato digital, lo que simplifica la logística de administrar estos instrumentos y se ajusta perfectamente a su formato presencial para crear condiciones favorables para la implementación. Además, factores externos como la pronta apertura de las escuelas (en 2020) y la alta tasa de la población objetivo con acceso a una computadora durante la pandemia (90 por ciento) también facilitaron la recolección (Herrero et al., 2022).

En El Salvador se creó la prueba AVANZO en formato digital, que reemplazó a la prueba PAES. Para 2020, la prueba podía ser tomada por el estudiante de forma flexible en un rango de tiempo determinado, cuando tuviera acceso a un dispositivo electrónico (MINEDUCYT, 2022). Después de la primera experiencia de la prueba en 2020, en 2021 se extendió una política que dotó a todos las y los estudiantes de último año de educación media de computadoras portátiles no retornables para poder presentar la prueba y continuar sus estudios superiores. Finalmente, en Colombia se introdujeron cambios en la aplicación del examen Saber 11 en 2020. Por ejemplo, se redujo el número de preguntas en cada prueba (excepto para el cuestionario socioeconómico) y las personas evaluadas respondieron el exa-

men en una sola sesión (normalmente eran dos sesiones) de cinco horas y treinta minutos, en aulas acondicionadas tanto para preservar su salud como para no ser fuente de contagio masivo entre los asistentes (ICFES, 2022). Aunque en 2020 la prueba no se realizó en formato digital (a diferencia de otras pruebas realizadas por la entidad como Saber Pro para educación universitaria y Saber T&T para educación técnica y tecnológica profesional), el instituto ha realizado varios pilotos digitales para su eventual presentación digital desde 2021.

Evaluaciones formativas y diagnósticas

Luego de la pandemia hay un renovado interés en la región por difundir, promover e institucionalizar la evaluación formativa y diagnóstica. A diferencia de la evaluación sumativa (o evaluación del aprendizaje), que se concentra en el final de una etapa educativa, la evaluación formativa (o evaluación para el aprendizaje) se realiza durante todo el proceso, proporcionando retroalimentación continua tanto al estudiante como al docente para ajustar y mejorar sus estrategias de aprendizaje y enseñanza (UNESCO, 2021). Las evaluaciones diagnósticas, que también tienen un enfoque formativo, tienen la particularidad de aplicarse al inicio de un ciclo o año lectivo, con el objetivo de identificar el punto de partida de los estudiantes y orientar los procesos de enseñanza y aprendizaje acorde a las necesidades de los estudiantes. Desde 2020 varios países implementaron nuevas evaluaciones formativas y/o diagnósticas para monitorear los aprendizajes (UNESCO, 2021). Honduras implementó por única vez las evaluaciones diagnósticas de los aprendizajes en 2020. Desde 2021 Chile ha utilizado la plataforma Diagnóstico Integral de Aprendizajes (DIA), de uso voluntario y disponible para todas las escuelas del país, para monitorear los aprendizajes académicos y socioemocionales de las y los estudiantes en tres momentos del año escolar. En 2021 Brasil lanzó las evaluaciones diagnósticas formativas para todos los grados de primaria de las escuelas públicas y desde

2022 también para secundaria. En 2021 Guatemala implementó la evaluación diagnóstica de estudiantes (ED) y en 2022 y 2023 las pruebas formativas Aprender + para todos los grados de primaria y secundaria. En 2022 México incluyó dentro de su modelo de evaluación un enfoque de evaluaciones diagnósticas formativas anuales de segundo a sexto grado de primaria y primero a tercer grado de secundaria en las áreas de lectura, matemáticas y formación cívica y ética con distintas alternativas de aplicación (en línea, sistema digital o papel y lápiz). El Salvador recolectó entre 2015 y 2018 la Evaluación de Capacidades Productivas y Ciudadanas (ECPYC) y desde 2021 las evaluaciones Conociendo mis logros desde tercer grado de primaria hasta segundo año de bachillerato; estas evaluaciones tenían fines reflexivos y de implementación de planes de refuerzo y mejora continua. Finalmente, desde 2011 Uruguay pone anualmente a disposición de los docentes evaluaciones formativas para que las puedan aplicar a sus estudiantes de educación primaria y secundaria, a través de la plataforma SEA; los docentes pueden allí incluso crear su propia prueba a partir de las actividades liberadas del banco de ítems, para aplicar a sus estudiantes evaluaciones formativas cuando lo consideren conveniente.

Si bien la mayoría de los países ha adoptado las pruebas diagnósticas y/o formativas como complemento de las pruebas nacionales estandarizadas, países como El Salvador, México y Guatemala no han retomado las evaluaciones nacionales sumativas a gran escala para primaria después de 2020, enfocándose únicamente en pruebas diagnósticas y/o formativas como instrumento de evaluación para los aprendizajes.

II.3 Reflexiones finales: hacia una cultura de evaluación educativa

Como hemos visto, la evolución de los sistemas de evaluación educativa de América latina y el Caribe ha sido muy variada, pero los sistemas comparten muchas similitudes en términos de los desafíos que enfrentan. Contar con herramientas sólidas para el monitoreo de los aprendizajes sigue siendo un objetivo claro de la política pública educativa en los países de la región. El debate a futuro se centra en el contenido, las características y los usos que tendrán las pruebas de aprendizaje para dar respuestas a los retos de calidad que persisten en la región. A continuación describimos los principales temas que están marcando esas discusiones sobre la base del análisis realizado y de entrevistas con líderes de la evaluación educativa de algunos países de América Latina⁴².

Creación de capacidades. Los sistemas nacionales de evaluación de la región son heterogéneos en términos de su madurez y de la capacidad técnica de sus equipos, de la periodicidad, confiabilidad, trazabilidad y comparabilidad de sus pruebas y del uso de resultados. En este sentido, una prueba regional como ERCE no sólo permite a los países cumplir mínimamente con el objetivo de monitoreo del ODS 4; además ofrece a los países interesados en formar capacidades técnicas e ir construyendo sistemas de evaluación sólidos una oportunidad para capacitar a sus equipos técnicos. Efectivamente, participar de las pruebas internacionales significa la posibilidad de aprovechar espacios favorables para la formación de capacidades técnicas, aun cuando requieran inversiones adicionales. Ante la ausencia o interrupción de las evaluaciones nacionales, las evaluaciones internacionales permiten observar el estado de avance de los sistemas educativos; por ejemplo, en el monitoreo del ODS 4. Por ello las evaluaciones internacionales juegan un papel importante en la región y debe fortalecerse su continuidad. Es crucial resaltar la carencia de centros

académicos especializados en la medición de la calidad educativa en la región. En este contexto, Chile y Brasil se destacan con centros como el Centro de Medición MIDE UC y el Centro de Políticas Públicas e Avaliação da Educação (CAEd); ambas han sido clave para el desarrollo y fortalecimiento de los sistemas de evaluación educativa, tanto a nivel nacional como regional. La escasez de expertos en psicometría en América Latina subraya la necesidad de invertir en centros especializados y programas de doctorado en esta área, lo que mejoraría significativamente la capacidad técnica y la calidad de los sistemas de evaluación en la región.

La eficiencia en la entrega y uso de resultados es uno de los grandes desafíos de los sistemas educativos. El procesamiento de los resultados de las pruebas, la entrega y divulgación de informes de resultados y su utilización en los centros educativos –y más aún para la población docente de forma que redunden en mejoras en el salón de clase– es un gran desafío y casi una utopía. Hay sistemas educativos que han logrado minimizar los tiempos de procesamiento de resultados para publicar los resultados al menos durante el mismo año escolar, pero esto sucede en pocos casos. La apropiación y uso de los resultados debe ser uno de los objetivos principales de las evaluaciones de aprendizaje y en ese sentido las entidades a cargo deben orientar sus esfuerzos para generar informes de resultados pertinentes y en el menor tiempo posible.

Balance del uso de los distintos tipos de pruebas (diagnósticas, formativas y sumativas) para la mejora de los aprendizajes. Es saludable que en los sistemas de evaluación convivan diferentes tipos de evaluaciones para instaurar un sistema de pesos y contrapesos que permita brindar información sobre el desempeño de las y los estudiantes a las y los responsables de la política educativa, autoridades educativas, a los centros educativos, a la población docente en los salones de clase

42 Las personas entrevistadas fueron, en orden alfabético, las siguientes: Gina Garcés, Directora de Evaluación, Ministerio de Educación de Panamá; Carlos Henríquez, Coordinador de LLECE y exdirector de la Agencia de Calidad de la Educación de Chile; Harvey Sánchez, CEO de la Agencia Latinoamericana de Evaluación y Política Pública y Director Ejecutivo de INEVAL en Ecuador; y Ansell Schecker, viceministra de Asuntos Técnicos y Pedagógicos del Ministerio de Educación de República Dominicana.

y a las familias y cuidadores. Esa información proviene principalmente de evaluaciones de aprendizaje y de cuestionarios de factores asociados. Dentro de las diferentes evaluaciones para el aprendizaje, las pruebas diagnósticas y formativas se caracterizan por ser de aplicación por la población docente, con cuestionarios liberados y con entrega muy rápida de resultados comparados. Por el contrario, las evaluaciones sumativas de gran escala, en cambio, tienen características técnicas más rigurosas para garantizar confiabilidad, comparabilidad y trazabilidad. Las evaluaciones formativas requieren una apropiación o un profundo conocimiento de los objetivos de aprendizaje por parte de la población docente. Para llegar a ese punto se necesita un grado de madurez alto en conocimientos de evaluación. Quizás por ello los sistemas educativos han recorrido un camino natural que empieza por las evaluaciones estandarizadas a gran escala y a través de procesos de divulgación de resultados, que incluyen formación sobre elaboración de ítems, interpretación de resultados y recomendaciones pedagógicas. La población docente se apropia del uso de resultados que le permite usar evaluaciones formativas, administradas directamente con una entrega de resultados en menor tiempo y con intervenciones más rápidas en el salón de clase. Este es un ejemplo concreto de cómo ambos tipos de pruebas pueden convivir con ciertas precauciones: con el cronograma de aplicación de las pruebas, para no tener aplicaciones simultáneas que generen interrupciones constantes en el calendario escolar; en la aplicación a diferentes grados para evitar tener poblaciones sobreevaluadas y otras sin evaluaciones; y sobre todo poniendo el foco en la entrega de reportes de resultados orientados para el uso en el salón de clase de tal forma que la población docente pueda utilizarlos para implementar acciones de mejoramiento a la mayor brevedad.

La financiación de las evaluaciones nacionales depende principalmente de una asignación presupuestaria anual importante como porcentaje del presupuesto de las carteras de educación. Por eso ante escenarios de menores recursos en educación las autoridades deciden postergar o no aplicar las evaluaciones, afectando el cronograma de evaluación y el monitoreo de los indicadores de desempeño de los y las estudiantes en el sistema educativo. Dada la importancia para las autoridades educativas de tomar decisiones basadas en información, es importante garantizar el cumplimiento del cronograma de aplicación de evaluaciones, ya que las evaluaciones son la fuente principal de información para la toma de decisiones.

Calidad y validez. La consolidación de los sistemas nacionales de evaluación requiere un plan de pruebas sostenible, con pruebas técnicamente confiables y comparables. Cuando ese plan logra implementarse y se corrobora la calidad y validez de las pruebas, es posible afirmar que el sistema de evaluación ha alcanzado un alto grado de madurez. Desafortunadamente, pocos programas de evaluación educativa han acumulado suficiente evidencia para respaldar sus interpretaciones y usos. La ausencia de evidencia de validez representa una seria amenaza para las evaluaciones, comprometiendo su valor y su viabilidad política y técnica. La consolidación de los sistemas de evaluación en la región se ha visto afectada por la falta de cumplimiento de los planes de evaluación, por los cambios en las líneas de base o criterios de evaluación, por falta de financiamiento y de recursos humanos con formación técnica o por falta de voluntad política. El compromiso con y la adherencia a los planes de evaluación puede ser apoyado desde la comunidad educativa cuando hay confianza en los sistemas de evaluación y cuando la comunidad percibe la utilidad de los resultados. Adicionalmente, organismos multilaterales que en ocasiones apoyan el financiamiento de las pruebas o agencias internacionales que promueven el fortalecimiento de capacidades pueden hacer *advocacy* para mantener los sistemas de evaluación.

Medición generalizada de componentes socioemocionales. La reapertura de las escuelas tras el prolongado cierre por la pandemia trajo consigo una discusión profunda sobre cómo acoger a niños y niñas, jóvenes, adolescentes y maestros y maestras. A pesar de la premura por identificar el rezago en el logro de los aprendizajes, las afectaciones socioemocionales debieron ser atendidas prioritariamente para restaurar el clima escolar previo al cierre de las escuelas. En ese sentido, también es una prioridad medir los componentes socioemocionales para identificar intervenciones de recuperación conducentes a restablecer ambientes de enseñanza-aprendizaje. Además, la pandemia impulsó otro tipo de cambios en los componentes de las pruebas. En Ecuador se aplicó un módulo para obtener información sobre la situación de las y los estudiantes durante la pandemia y se incluyó un constructo socioemocional. En un estudio longitudinal promovido desde la sociedad civil y el BID, en Brasil se aplicaron nueve olas de encuestas dirigidas a familias durante la pandemia y durante el proceso de reapertura de escuelas para estudiar la percepción de las familias en este proceso; esos estudios fueron replicados en Colombia y Uruguay. En Chile y Perú también hubo un gran foco en monitorear los componentes socioemocionales durante y después del cierre de escuelas.

El uso de la tecnología ha sido ampliamente discutido como una alternativa costo-eficiente para los países. La revisión de experiencias en pruebas por computadora y su potencial implementación en el futuro dejó en claro que los países de la región aún no cuentan con la infraestructura y la conectividad necesarias para un despliegue de pruebas censales por computadora; su implementación requiere una cobertura de 100 por ciento que aún no se cumple. La alternativa para poder cubrir todo el territorio sería adquirir computadoras para las aplicaciones o alquilarlas y movilizarlas por el país, lo cual incrementa los costos y no resulta eficiente. En cuanto esa barrera no sea superada, la posibilidad de moverse en esa dirección aún es lejana. Esto no debe compararse con pruebas muestrales por computadora, que tie-

nen un esquema diferente dado el tamaño de la muestra y el despliegue logístico que debe implementarse; esto las hace viables y amerita estudios comparativos de costo-efectividad entre las dos modalidades, pues los tiempos de procesamiento y entrega de resultados podrían reducirse significativamente.

Aún tenemos un largo camino por recorrer para establecer una verdadera cultura de evaluación en los sistemas educativos de la región. En el contexto postpandemia, las evaluaciones constituyen un instrumento esencial para poder identificar cómo fueron afectados los sistemas educativos y utilizar los resultados para diseñar estrategias de recuperación de aprendizajes. Sin embargo, la interrupción en los calendarios de aplicación de las evaluaciones y su lenta recuperación a los niveles de medición de 2019 podría ser un indicio de la resistencia que aún persiste frente a aplicaciones que dejan en evidencia menores desempeños en comparación con mediciones anteriores. Los sistemas educativos perdieron varios meses (o años) valiosos en ese proceso, pero aún estamos a tiempo de retomar y seguir construyendo nuestros sistemas nacionales de evaluación.

Anexo

Metodología del informe

Los datos utilizados para este estudio fueron recopilados entre septiembre de 2023 y marzo de 2024. La metodología se basó en gran medida en la búsqueda y revisión de documentos públicos y comunicaciones oficiales de las autoridades educativas y entidades de evaluación y entrevistas focalizadas con expertos que han tenido a cargo la dirección de los departamentos o instituciones de evaluación educativa en Ecuador, Panamá, República Dominicana y Chile.

Las evaluaciones que se tomaron en cuenta para cada país del análisis comprenden prue-

bas de aprendizaje a gran escala sumativas levantadas por 18 países de América Latina entre 1990 y 2023 en los niveles de educación primaria y secundaria. Entre las evaluaciones mencionadas, el informe se centra en evaluaciones con fines de monitoreo, cuyas implicaciones pueden ser altas o bajas; y exámenes para la certificación de un curso o nivel educativo (primario o secundario) o para la selección o admisión a la educación superior siempre y cuando este forme parte de la malla curricular de la educación secundaria o primaria.

Las evaluaciones analizadas se detallan en la siguiente tabla.

País	Nombre de la(s) prueba(s)
Argentina	Operativo Nacional de Evaluación ONE, Aprender
Bolivia	Sistema de Medición y Evaluación de la Calidad de la Educación SIMECAL
Brasil	SAEB, Aneb, Anresc, ANA
Chile	Sistema nacional de evaluación de resultados de aprendizaje SIMCE
Colombia	Pruebas SABER 11, 9, 7, 5, 3
Costa Rica	Pruebas Nacionales Diagnósticas MEP, Bachillerato, III Ciclo, FARO
Ecuador	Aprendo, SER
El Salvador	Prueba de Acceso a la Educación Superior PAES, Avanzo
Guatemala	Evaluación del Rendimiento Escolar PRONERE, Evaluación a Tercero Básico TER
Honduras	Evaluaciones Rendimiento Académico
México	Estándares Nacionales, Enlace, Exámenes de la Calidad y el Logro Educativos (EXCALE), PLANEA
Nicaragua	Sistema Nacional de Evaluación de la Educación Básica y Media, Evaluación Nacional
Panamá	Sistema Nacional de Evaluación de la Calidad SINECA, Crecer
Paraguay	SNEPE, Sistema Nacional de Evaluación del Proceso Educativo, ENLACE
Perú	Evaluación Muestral de Estudiantes, Evaluación Censal de Estudiantes
República Dominicana	Sistema de Pruebas Nacionales, pruebas diagnósticas
Uruguay	Evaluación Nacional de Aprendizajes, Aristas
Venezuela	SINEA, Sistema Nacional de Medición y Evaluación del Aprendizaje

Referencias

- Altinok, N., Angrist, N. y Patrinos, H. A.** (2018). Global data set on education quality (1965-2015). *World Bank Policy Research Working Paper*, (8314).
- Arias Ortiz, E., Bos, M. S., Giambruno, C. y Zoido, P.** (2023a). *América Latina y el Caribe en PISA 2022: ¿cuántos tienen bajo desempeño?* CIMA, Banco Interamericano de Desarrollo. <https://doi.org/10.18235/0005316>
- Arias Ortiz, E., Bos, M. M. S., Giambruno, C. y Zoido, P.** (2023b). *América Latina y el Caribe en PISA 2022: ¿cómo le fue a la región?* CIMA, Banco Interamericano de Desarrollo. <https://doi.org/10.18235/0005318>
- Arias Ortiz, E., Giambruno, C., Morduchowicz, A. y Pineda, B.** (2023c). *El estado de la educación en América Latina y el Caribe 2023*. Banco Interamericano de Desarrollo. <https://doi.org/10.18235/0005515>
- Arias Ortiz, E., Bos, M. S., Chen Peraza, J., Giambruno, C., Levin, V., Oubiña, V., Pineda, J. A. y Zoido, P.** (2024). *El aprendizaje no puede esperar: Lecciones para América Latina y el Caribe a partir de PISA 2022*. Banco Interamericano de Desarrollo y Banco Mundial. <https://doi.org/10.18235/0005671>
- Arias, E., Hincapié, D. y Paredes, D.** (2020). *Educación para la vida: El desarrollo de las habilidades socioemocionales y el rol de los docentes*. Banco Interamericano de Desarrollo.
- Arregui, P.** (2008). *Uso de los resultados de evaluaciones educativas a gran escala en América Latina*. PREAL. Grupo de Trabajo sobre Estándares y Evaluación.
- Asamblea General de las Naciones Unidas.** (2015). *Resolución aprobada por la Asamblea General el 25 de septiembre de 2015 [sin remisión previa a una Comisión Principal (A/70/L.1)] (Documento oficial No. 70/1). Transformar nuestro mundo: la Agenda 2030 para el Desarrollo Sostenible*. Recuperado de https://unctad.org/system/files/official-document/ares70d1_es.pdf
- Ascorra, P., Carrasco, C., López, V. y Morales, M.** (2019). *Políticas de convivencia escolar en tiempos de rendición de cuentas*. Education Policy Analysis Archives, 1-24.
- Balarin, M.** (2021). *The Political Economy of Education Reforms in Peru 1995-2020*. GRADE.
- Banco Mundial.** (2018). *Using student assessments to improve instruction and learning*. World Bank.
- Blazer, C.** (2011). *Unintended Consequences of High-stakes Testing*. Information Capsule.
- Borger, M., Elacqua, G., Jacas, I., Neilson, C. y Olsen, A. S. W.** (2024). Report cards: Parental preferences, information and school choice in Haiti. *Economics of Education Review*, 102, 102560.
- Brooke, N.** (2016). High-stakes accountability using teacher salary incentives in Brazil: An update. *Revista de Currículum y Formación de Profesorado*, 207-250.
- Bruns, B., Akmal, M. y Birdsall, N.** (2019). *The Political Economy of Testing in Latin America and Sub-Saharan Africa*. RISE.
- Cáceres, C., Muñoz, C. y Valenzuela, J.** (2021). *Responsabilidad personal docente y motivación escolar*. *Revista Electrónica Interuniversitaria de Formación del Profesorado*, 175-188.
- Campos Morán, S., Navarrete Gálvez, P. M. y Campos Solórzano, M. R.** (2023). *Evaluación estandarizada de competencias en El Salvador: Legado del modelo PAES y perspectivas de la prueba AVANZO*. Entorno, 50-59.
- Caribbean Examinations Council.** (s.f.). Caribbean Examinations Council. <https://www.cxc.org/about/about-cxc/>
- Caribbean Examinations Council** (2023). Caribbean Examinations Council: Annual Report 2022. Disponible en: <https://www.cxc.org/cxc-annual-report-2022/>

Cariola, L., Covacevich, C., Grubler, J., Lagos, E. y Ortiz, M. (2011). *Chilean participation in IEA studies. In IEA 1958-2008: 50 years of experiences and memories* (Vol. 2008, p. 50).

Centro de Información para la Mejora de los Aprendizajes - CIMA. (n.d.). *Banco Interamericano de Desarrollo*. <https://cima.iadb.org/es>

Clarke, M. y Lune-Bazaldúa, D. (2021). *Primer on Large-Scale Assessments of Educational Achievement*. World Bank.

Consejo Nacional de Rectores. (2023). *Noveno Estado de la Educación 2023*. Programa Estado de la Nación.

Cox, C. y Meckes, L. (2016). International large-scale assessment studies and educational policy-making in Chile: contexts and dimensions of influence. *Research Papers in Education*, 31(5), 502-515.

Cueto, S. (2007). Las evaluaciones nacionales e internacionales de rendimiento escolar en el Perú: balance y perspectivas. En *Investigación, políticas y desarrollo en el Perú* (pp. 405-455). GRADE.

De Hoyos, R., Ganimian, A. y Holland, P. (2019). *Teaching with the Test: Experimental Evidence on Diagnostic Feedback and Capacity-Building for Schools in Argentina*. World Bank, 123-140.

Dueñas, X., Elacqua, G., Jaramillo, L. y Margitic, J. F. (2022). *Pérdidas de aprendizaje debido al COVID-19. Desafíos y hallazgos en Colombia*. BID.

Elacqua, G., Martínez, M. y Westh, A. S. (2019). *Diseño de índices de calidad escolar: Lecciones de la experiencia internacional*. BID.

Elacqua, G. y Fabrega, R. (2004). *El consumidor de la educación: El actor olvidado de la libre elección de escuelas en Chile*. Santiago.

FEREMA. (2022). *Progreso Educativo de Honduras (IPEH) 2022*. FEREMA.

Fernandes, R. y Gremaud, A. P. (2020). Large-scale educational assessment and accountability: a brief analysis of the Brazilian experience.

Ferrer, G. (2006). *Educational Assessment Systems in Latin America: Current Practice and Future Challenges*. Partnership for Educational Revitalization in the Americas (PREAL).

Ferrer, G. y Fiszbein, A. (2015). *¿Qué ha sucedido con los sistemas de evaluación de aprendizajes en América Latina?* Washington DC: El Diálogo.

Flamand, L., Arriaga, R. y Santizo, C. (2020). *Reforma educativa y políticas de evaluación en México. ¿Instrumentos para abatir el rezago escolar y promover la igualdad de oportunidades?* *Revista Foro Internacional (Foro Internacional FI)*, 717-753.

Galas, C., Gutiérrez, M. T. y Hamilton, M. E. (2020). *Uso de resultados de las pruebas de aprendizaje en el diseño de las políticas educativas en América Latina*. UNESCO, Buenos Aires.

Gordon, E. D. (2019). *Problems, pressures and policies affecting the progress of the Caribbean Examinations Council examinations: A postcolonial response to secondary education in Jamaica* (Doctoral dissertation, UCL (University College London)).

Harris, A. (1991). *Effective Assessment of Educational Progress: A Review of Strategies for Measuring Learning Achievement*. World Bank.

Herrero Tejada, A., López Leavy, M., Finoli, M., Hevia de la Jara, F., Stanton, S. y Fiszbein, A. (2022). *Evaluación en pandemia: ¿Cómo diagnosticamos las pérdidas de aprendizajes para informar los esfuerzos de recuperación?* Diálogo Interamericano

Heyneman, S. P. y Lee, B. (2014). The impact of international studies of academic achievement on policy and research. *Handbook of international large-scale assessment: Background, technical issues and methods of data analysis*, 37-72.

Instituto Colombiano para la Evaluación de la Educación - ICFES. (2017). *Experiencia acumulada en el levantamiento de datos de factores asociados, 2012-2015 SABER 3, 5 y 9*. ICFES, Bogotá.

Instituto Colombiano para la Evaluación de la Educación - ICFES. (2021). *Habilidades Socioemocionales. Marco de referencia para la evaluación ICFES.* ICFES, Bogotá.

Instituto Colombiano para la Evaluación de la Educación - ICFES. (2022). *Informe nacional de resultados del examen Saber 11° 2020.* ICFES, Bogotá.

Keeves, J. P. (1995). *The world of school learning: selected key findings from 35 years of IEA research.* International association for the evaluation of educational achievement (IEA).

Martínez, F. (2015). *Las pruebas ENLACE y EXCALE, un estudio de validación.* INEE, Ciudad de México.

Martínez-Iñiguez, J., Tobón, S., López-Ramírez, E. y Manzanilla-Granados, H. M. (2020). *Calidad educativa: un estudio documental desde una perspectiva socioformativa.* *Revista Latinoamericana de Estudios Educativos*, 233-258.

Ministerio de Educación - MINEDUCA. (2014). *Hacia un sistema completo y equilibrado de evaluación de los aprendizajes en Chile. Informe Equipo de Tarea para la Revisión del SIMCE.* MINEDUCA, Santiago.

Ministerio de Educación Pública - MEP. (2022). *Informe Nacional Pruebas Nacionales FARO: Primaria – Secundaria, II semestre-2021.* Ministerio de Educación Pública, San José.

Ministerio de Educación, Ciencia y Tecnología - MINEDUCYT. (2022). *Informe de Resultados Avanzo.* MINEDUCYT, San Salvador.

Montoya, S. (2023). Reporting learning outcomes in basic education: Country's options for indicator 4.1.1. *UNESCO Institute for Statistics. Montreal, Canada.* https://tcg.uis.unesco.org/wp-content/uploads/sites/4/2022/08/Countrys-reporting-option-_Zambia_AAEA.Final_.pdf

Murillo, F. J. (2007). Resultados de aprendizaje en América Latina a partir de las evaluaciones nacionales. UNESCO/OREALC.

OCDE, UNESCO, UNICEF y Banco Mundial. (2022). *From Learning Recovery to Education Transformation: Insights and Reflections from the 4th Survey on National Education Responses to COVID-19 School Closures.* OCDE, UNESCO, UNICEF, Banco Mundial, París, Montreal, New York, Washington D. C.

OCDE. (2018). *PISA for development assessment and analytical framework: Reading, mathematics and science.* OECD Publishing.

OCDE. (2023). *PISA 2022 Results (Volume I): The State of Learning and Equity in Education.* OECD Publishing. <https://doi.org/10.1787/53f23881-en>

OCDE. (2021). *National Assessment Reform: Core Considerations for Brazil.* OECD Education Policy Perspectives, París.

Pérez, G. y Maldonado, C. (2015). *Panorama de los sistemas de monitoreo y evaluación en América Latina.* CIDE, México.

Pérez, G. y Serrano, S. (2016). *Impacto de la evaluación estandarizada en los sistemas educativos en América Latina.* Banco Interamericano de Desarrollo.

Ramirez, F. O., Schofer, E. y Meyer, J.W. (2018). "International Tests, National Assessments, and Educational Development (1970–2012)". *Comparative Education Review*, vol. 62, núm. 3, págs. 344 a 364.

Rivas, A. (2015). *América Latina después de PISA: Lecciones aprendidas de la educación en siete países (2000-2015).* Fundación Cippec.

Rodríguez-Garcés, C., Padilla-Fuentes, G. y Suazo Ruiz, C. (2020). *Medición de calidad educativa en Chile: Lo que reportan los indicadores de desarrollo cognitivo, personal y social en la escuela.* *Revista Pilquen - Sección Psicopedagogía*, 34-48.

Sánchez, M. y del Sagrario, F. (2013). Las evaluaciones estandarizadas: Sus efectos en tres países latinoamericanos. *Revista Latinoamericana de Estudios Educativos (México)*, 97-124.

Saracho Martínez, A. (2007). *Casos de Estudio de Reacciones a PISA 2000 y 2003: Alemania, Polonia, Brasil y Otros*. [Case Study of Reactions to PISA 2000 and 2003: Germany, Poland, Brazil and Others]. México, DF: Fundación IDEA.

Teig, N. y Steinmann, I. (2023). Leveraging large-scale assessments for effective and equitable school practices: The case of the Nordic countries. *Large-scale Assessments in Education*, 21.

Thailinger, A., Pecha, C., Beuermann, D., Ortiz, E. A., Hobbs, C. y Piras, C. (2023). Gender Gaps in the English-speaking Caribbean: Education, Skills, and Wages. InterAmerican Development Bank.

Tobin, M., Nugroho, D. y Lietz, P. (2016). Large-scale assessments of students' learning and education policy: synthesising evidence across world regions. *Research Papers in Education*, 31(5), 578-594.

UNESCO. (2019). La promesa de las evaluaciones de aprendizaje a gran escala. Reconocer los límites para generar oportunidades. UNESCO.

UNESCO. (2021). *Evaluación formativa: Una oportunidad para transformar la educación en tiempos de pandemia*. UNESCO.

UNESCO Institute for Statistics. (2024). GAML/TCG criteria for use of an assessment to report on SDG 4.1. 1. <https://tcg.uis.unesco.org/wp-content/uploads/sites/4/2024/02/GAML-Criteria-for-reporting.pdf>

UNESCO Institute for Statistics. (2013). *Clasificación Internacional Normalizada de la Educación CINE 2011*. UNESCO-UIS.

UNESCO Institute for Statistics. (2022). Metadata for Indicator 1.a.1 - Education expenditures per capita as a percentage of GDP. https://tcg.uis.unesco.org/wp-content/uploads/sites/4/2021/09/Metadata-FFA-1.a.gdp_.pdf

Vanni, X. y Valenzuela, J. P. B. (2020). Evaluación del Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación. *Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura*.

Vázquez Olivera, M. G. (2015). La calidad de la educación. Reformas educativas y control social en América Latina. *Revista de Estudios Latinoamericanos*, 93-124.

Vegas, E. y Petrow, J. (2008). *Learning in Latin America: The challenge for the 21st century*. Banco Mundial.

Wagemaker, H. (2014). International large-scale assessments: From research to policy. *Handbook of international large-scale assessment: Background, technical issues, and methods of data analysis*, 11-36.

Zea, M. (2020). La evaluación de la calidad educativa en el paradigma de la educación para vivir en Bolivia. *Journal of Supranational Policies of Education*, 106-126.

2024

EL ESTADO DE LA
EDUCACIÓN
en América Latina y el Caribe

LA MEDICIÓN DE LOS APRENDIZAJES

Elena Arias Ortiz, Ximena Dueñas, Cecilia Giambruno y Angela López

