



---

## Noveno Informe Estado de la Educación (2023)

---

### Investigación

---

Aspectos claves a considerar en el desarrollo de un sistema nacional público de macroevaluación de los aprendizajes de los estudiantes de Costa Rica: recomendaciones desde la experiencia peruana

**Investigadora:**

Lilliana Miranda Molina

San José | 2023



370.7286  
M672a

Miranda Molina, Lilliana

Aspectos claves a considerar en el desarrollo de un sistema nacional público de macroevaluación de los aprendizajes de los estudiantes de Costa Rica: recomendaciones desde la experiencia peruana / Lilliana Miranda Molina -- Datos electrónicos (1 archivo : 677 kb). -- San José, C.R. : CONARE - PEN, 2023.

ISBN 978-9930-618-68-4

Formato PDF, 36 páginas.

Investigación para el Noveno Informe Estado de la Educación 2023

1. EDUCACIÓN. 2. ESTUDIANTES. 3. APRENDIZAJE. 4. SISTEMA EDUCATIVO. 5. EVALUACIÓN EDUCATIVA. 6. POLÍTICA EDUCATIVA. I. Título.



## Indice

Resumen .....	4
Modelo de evaluación .....	6
Gobernanza: organización institucional .....	7
Gobernanza: Propósitos .....	11
Las evaluaciones sumativas .....	13
Evaluaciones formativas .....	13
Técnicas e instrumentos de medición.....	18
Referentes conceptuales para la evaluación: alineamiento curricular y progresiones de aprendizaje.....	18
TIC: Hacia formas más innovadoras de evaluar.....	20
Factores asociados: presentar resultados comparables.....	21
Difusión y uso de resultados.....	22
Uso de resultados para mejorar la práctica pedagógica.....	23
Uso de resultados para el diseño e implementación de políticas educativas .....	28
Referencias .....	29
Anexos.....	31

## **Descargo de responsabilidad**

Esta Investigación se realizó para el *Noveno Informe Estado de la Educación (2023)*, con financiamiento de la Fundación Yamuni Tabush. El contenido de la ponencia es responsabilidad exclusiva de su autor, y las cifras pueden no coincidir con las consignadas en el *Noveno Informe Estado de la educación (2023)* en el capítulo respectivo, debido a revisiones posteriores. En caso de encontrarse diferencia entre ambas fuentes, prevalecen las publicadas en el Informe.

Formato aplicado por: Karol Arroyo Mora y revisado por: Giselle Rojas Arias.

## **Resumen**

El presente documento reúne un conjunto acotado de propuestas sobre el tema: “Aspectos claves a considerar en el desarrollo de un sistema nacional público de macro-evaluación de los aprendizajes de los estudiantes de Costa Rica: recomendaciones desde la experiencia peruana”.

Para su elaboración hemos buscado responder a la siguiente pregunta: ¿qué aspectos claves deben considerarse para establecer, en un plazo de un quinquenio, un sistema nacional público de macro-evaluación de los aprendizajes de los estudiantes robusto, transparente y efectivo que genere información de calidad para apoyar a los docentes y mejorar sus prácticas y mediación pedagógicas en las aulas y los aprendizajes de los estudiantes?

En el ámbito académico y de gestión educativa existe un amplio consenso acerca de la importancia fundamental de los propósitos que establece un sistema de evaluación, en tanto definen las características institucionales y técnicas del mismo. Por lo tanto, si bien el presente documento se ha ceñido a los Términos de Referencia de la consultoría y al documento de preguntas relevantes provisto por el Programa Estado Nación, en la sección 2 se encuentra el núcleo donde se desarrollan las ideas fuerza de las recomendaciones. Las secciones 3 (técnicas e instrumentos) y 4 (difusión y uso de resultados) se derivan de lo sugerido en la sección 2. Es decir, estas secciones abordan aspectos más puntuales y tienen un carácter más instrumental.

El documento tiene un carácter propositivo que privilegia el análisis crítico de la experiencia peruana en materia de sistemas de evaluación haciendo referencia documentada a las buenas prácticas nacionales, así como al conocimiento de experiencias de otros países. En este sentido, los criterios empleados en el análisis de la información y experiencias nos han permitido identificar una selección de recomendaciones apoyada en una justificación de las razones por las que fueron escogidas. La experiencia peruana puede considerarse relevante pero no exenta de problemas, errores y debates a lo largo de su desarrollo institucional. Por esta razón, el documento ha supuesto una mirada crítica al caso peruano en diálogo con otras experiencias latinoamericanas.

Al tener la oportunidad Costa Rica de repensar su sistema de macro evaluación debe tomar en cuenta la experiencia sobre lo que ya han recorrido los países que tienen una trayectoria importante en evaluaciones de sistema. Los países de la región llevan un recorrido de casi tres décadas de evaluación de aprendizajes y en ese trayecto han repensado muchas de sus decisiones iniciales. De otro lado, lo que se encuentra es que no existen sistemas de evaluación

que respondan únicamente a un modelo, casi todos han ido ajustando, flexibilizando y ampliando sus esquemas sin perder del todo la identidad con que fueron diseñados. En ese sentido, considerando que las evaluaciones de sistemas acarrear consecuencias no queridas. Incluso las evaluaciones de bajas consecuencias (Unesco, 2019), se ha propuesto para los distintos aspectos solicitados en la consultoría, algunas estrategias y mecanismos que amortigüen esos efectos.

La propuesta del sistema de evaluación comprende la formación de un instituto autónomo y técnico encargado del diseño y conducción del proceso de macro-evaluación en Costa Rica. Esta agencia especializada supone un nivel de autonomía frente a otras agencias y actores en el ámbito educativo. La autonomía institucional garantiza rigurosidad técnica, solvencia profesional y transparencia en los resultados de las evaluaciones. Sin embargo, puede derivar en alejamiento de las necesidades públicas e indiferencia frente a los debates de la política educativa. Por esta razón, la propuesta contempla algunas instancias orgánicas que vinculan al Instituto con el Ministerio de Educación y el Consejo Superior de Educación permitiendo contrapesos y canales institucionales de diálogo, control y rendición de cuentas.

De otro lado se sugiere que Costa Rica apueste por la conformación de un sistema de macro-evaluación comprensivo que incorpore a las evaluaciones sumativas, propias de las evaluaciones estandarizadas, dispositivos de evaluación formativa gestionados por los propios equipos de los centros educativos que permita tomar decisiones en los distintos niveles. Este diseño podría afianzar la legitimidad de la evaluación externa, aproximar a ésta con la evaluación de aula y ofrecer una plataforma desde la cual se puede establecer un diálogo con los docentes que aumente la probabilidad de que los resultados de las evaluaciones se aprovechen en todo su potencial. Esta apuesta también permite atender a las demandas que las áreas de evaluación han recibido para impactar en la transformación de las prácticas pedagógicas e institucionales antes que en el uso de la información que producen para el diseño e implementación de políticas educativas.

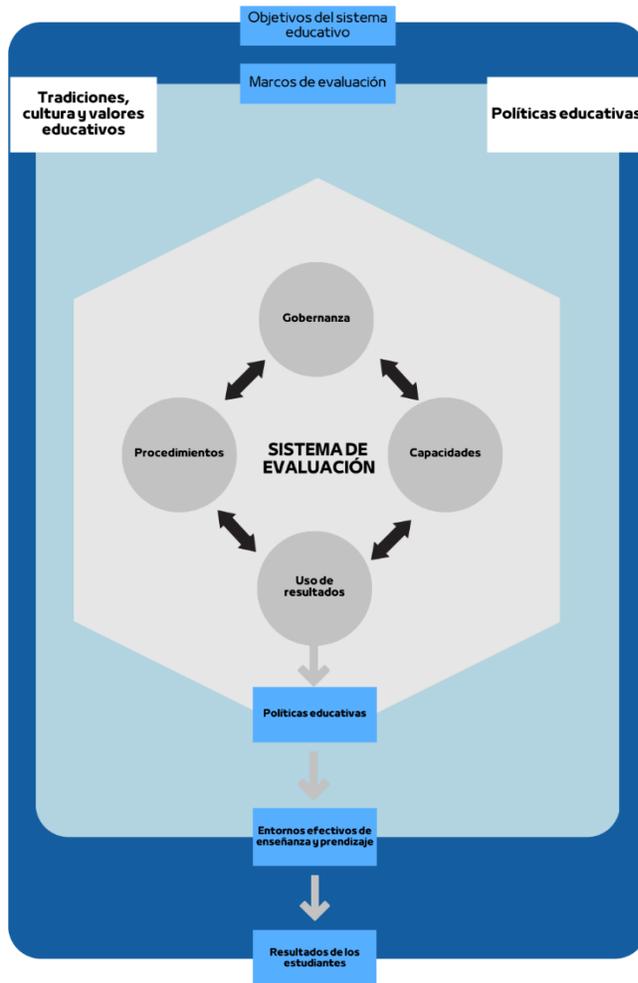
En dirección con lo anterior, se propone complementar la devolución de informes de resultados para distintos usuarios con la generación de espacios donde se coloque en un plano más amplio la lectura e interpretación de estos utilizando ya sea parte de los ítemes liberados o los mismos resultados. La dificultad de esta propuesta, ciertamente, es que la instancia de evaluación diseñe y ponga a disposición estos recursos, pero no se puede hacer cargo de su implementación, deben ser las áreas de formación docente o de currículo del Ministerio de Educación las que los incorporen en sus programas habituales de manera que se apunte a un cambio sistémico de las prácticas de aula.

Finalmente se recomienda que Costa Rica apueste por pruebas nacionales que tengan como referente el currículo nacional, pues ello permite orientar al docente en su práctica en el aula para lograr que los estudiantes logren los aprendizajes esperados y alinear y articular la política de evaluación de aprendizajes, la formación docente y la elaboración de materiales educativos. Asimismo, se sugiere que si deciden por elaborar estándares de aprendizaje evalúen el modelo de mapas de progreso dado que estos tienen un claro sentido pedagógico.

## Modelo de evaluación

Con el objetivo de que las propuestas del presente documento se entiendan mejor consideramos importante detallar el marco conceptual para las evaluaciones de logros de aprendizaje desarrollado por la OCDE (2013). Como se puede ver en el gráfico 1, ese marco involucra un conjunto de dimensiones interconectadas. Cabe señalar que a lo largo del documento solo nos referiremos a algunos de los aspectos que comprenden las dimensiones.

Gráfico 1  
Modelo de evaluación



Fuente: OCDE, 2013.

- **Gobernanza:** describe los diferentes propósitos del sistema de evaluación y el marco normativo para garantizar que los resultados sean utilizados en función a dichos objetivos. Esta dimensión también explora cómo están distribuidas las responsabilidades para implementar las evaluaciones y cómo los diferentes niveles de gobierno interactúan para conformar un sistema de evaluación coherente.

- **Procedimientos:** esta dimensión alude a los procedimientos y metodologías usadas para la evaluación. Esto incluye aspectos como las áreas de aprendizaje evaluadas, la selección de instrumentos y formatos, así como el uso de TIC. También comprende las consideraciones que se deben tomar en cuenta para garantizar una evaluación justa y equitativa.
- **Capacidad:** comprende las competencias y los apoyos necesarios para la evaluación y para aprovechar sus resultados. Incluye temas como: las capacidades que los estudiantes necesitan para participar y aprender de su evaluación; las competencias de evaluación que los profesores adquieren en la formación inicial docente y durante su desarrollo profesional y los arreglos y experiencia de las agencias involucradas en la evaluación de los estudiantes.
- **Uso de resultados:** esta dimensión incluye la forma en que se informan y utilizan los resultados de las evaluaciones con fines sumativos y formativos. Describe los estándares de calidad y los formatos de informes utilizados para los distintos usuarios. Asimismo, revisa los marcos legales vigentes para regular la manera cómo se difunden los resultados y analiza las formas en que los resultados de la evaluación se utilizan en diferentes contextos para proporcionar retroalimentación a los estudiantes y tomar decisiones respecto a su trayectoria educativa<sup>1</sup>.

## **Gobernanza: organización institucional**

Existen diferentes modelos institucionales para asumir las evaluaciones, pero siempre con presencia o representación del Estado. Así, están los modelos basados en el Ministerio de Educación, donde existe una oficina especializada en la evaluación del rendimiento o de la calidad educativa (por ejemplo, Argentina, Guatemala, Perú). Más recientemente se han empezado a crear organismos autónomos o semi autónomos especializados en evaluación (es el caso del INEP<sup>2</sup> de Brasil, ICFES<sup>3</sup> de Colombia, Ineval<sup>4</sup> de Ecuador). Finalmente, el caso de Uruguay se puede caracterizar como un formato mixto que combina un instituto nacional (INEED<sup>5</sup>) y la unidad de evaluación dentro de la Administración Nacional de Educación Pública (ANEP).

Cada modelo tiene pros y contras. Las oficinas dentro del Ministerio podrían facilitar la comunicación con otras unidades de la misma dependencia, que de hecho deberían ser algunos de los principales interesados. Un punto en contra de este arreglo es que mantener la oficina dentro del Ministerio la puede hacer más proclive a la influencia de presiones políticas, que en algunos casos se pueden oponer a la transparencia de la información. Las oficinas públicas autónomas deberían tener relativa mayor autonomía política pero justamente por esto sus resultados podrían tener menor impacto en el desarrollo de políticas educativas. De otro lado, el

---

<sup>1</sup> En el Anexo 1 se detallan las orientaciones de política para los sistemas de evaluación.

<sup>2</sup> <https://www.gov.br/inep/pt-br>

<sup>3</sup> <https://www.icfes.gov.co/>

<sup>4</sup> <https://www.evaluacion.gob.ec/>

<sup>5</sup> <https://www.ineed.edu.uy/>

formato mixto presenta en algunos casos conflictos por la superposición de funciones, pero también por la orientación evaluativa que privilegia cada instancia (Pinkasz, 2021).

En ese marco, se propone la formación de una agencia especializada (en adelante el “Instituto”) autónoma de las autoridades del Ministerio de Educación. El instituto deberá encargarse del diseño y conducción del proceso de macro-evaluación de Costa Rica.

Las razones que justifican esta propuesta son:

- Independencia frente a intervenciones del poder político: El instituto se crea con un marco jurídico de mayor rango (Ley General o reforma constitucional) que le permite disponer de la autoridad para mantener sus roles y competencias fuera de las orientaciones particulares o los objetivos de corto plazo promovidos por los gobiernos de turno.
- Autonomía técnica frente a inestabilidad política: Los cambios de gobierno y las crisis políticas generan tensiones en el sector educación pues los resultados de las evaluaciones pueden incrementar la deslegitimidad de un gobierno o de una gestión ministerial. En estos contextos, las presiones políticas se orientan a intervenir en el diseño de las pruebas o la publicación de sus resultados. Un instituto compuesto por cuadros técnicos solventes resulta más inmune a dichas interferencias políticas.
- Capacidad de planificar a mediano y largo plazo: Un instituto dispone de diversos mecanismos para garantizar una visión de mediano y largo plazo en materia de evaluación. La independencia frente al Ministerio de Educación y a los funcionarios que están orientados a proveer de resultados en el corto plazo al gobierno de turno, permite que el Instituto pueda construir una mirada estable de mediano y largo plazo.
- Capacidad para construir equipos técnicos solventes y estables: El instituto al contar con autonomía financiera y presupuestaria estará en mejores condiciones para tomar decisiones a largo plazo sobre los programas de evaluación (incluyendo la población a evaluar, la cobertura de contenido, los ciclos de administración de las pruebas, etc.) y mantener estos criterios y las decisiones que de ellos se desprenden durante un plazo de tiempo útil y pertinente.
- Sinergia y articulación transparente con el Ministerio de Educación: La autonomía puede convertirse en distanciamiento frente a los objetivos y preocupaciones del Ministerio de Educación, más aún cuando del distanciamiento se transita hacia la duplicidad de objetivos. Por esta razón en el diseño del Instituto debe considerarse la conformación de *vasos comunicantes* con el Ministerio de Educación, tales como la participación de un funcionario de alto rango en el Consejo Directivo del Instituto o reuniones periódicas de los equipos técnicos del instituto y el Ministerio de educación.
- Por último, nos parece importante señalar que, si se decide reanudar la aplicación de la Prueba Nacional de Bachillerato u otras pruebas con consecuencias, la experiencia internacional revela que los sistemas que tienen consecuencias en los examinados tienden a ser altamente sensible ante la opinión pública. Por ello, resulta recomendable diseñar y establecer una institucionalidad que simultáneamente pueda abordar las diversas aristas técnicas, comunicacionales y políticas en este ámbito, y que al mismo tiempo entregue

suficientes garantías de transparencia y equidad en sus decisiones. Esta sería una razón adicional para optar por la ruta de formación del Instituto.

Como evidencia empírica de lo señalado podemos mencionar que tres de los países con sistemas de evaluación más reconocidos de la región (Brasil, Chile, y Colombia) han formado institutos de evaluación adscritos al Estado, pero con algún nivel de autonomía respecto del Ministerio de Educación.

El modelo de Instituto garantiza la autonomía necesaria para que los procesos evaluativos y sus resultados sean independientes y confiables. Pero, como la evidencia empírica señala, la capacidad para informar verazmente a las autoridades y a la sociedad acerca de los resultados de las evaluaciones supone el desarrollo de un complejo manejo político para que los cambios en las coyunturas políticas no coloquen al Instituto como el responsable de tensiones y problemas no resueltos dentro de los sistemas educativos. Lo cual puede poner en peligro no solamente la autonomía, sino la existencia misma del Instituto y del sistema de evaluación en su conjunto. Tal fue el caso del Instituto Nacional para la Evaluación de la Educación (INEE) de México<sup>6</sup>.

La alternativa para evitar estos peligrosos escenarios es mantener vasos comunicantes entre el instituto y el Ministerio de Educación. Estos lazos pueden tener la forma de competencias específicas para el Ministerio de Educación en materia de evaluaciones educativas que complementen las funciones del Instituto; disposiciones normativas en la conformación de las instancias de gobierno del Instituto; representación directa de funcionarios ministeriales en el Instituto; entre otras.

De manera más específica y concreta recomendamos las siguientes medidas:

- El Directorio es el cuerpo directivo del Instituto. Funcionarios de la Alta Dirección del Ministerio de Educación debieran ser parte del Directorio del Instituto. Este órgano toma las decisiones fundamentales del sistema del sistema de evaluación (por ejemplo, aprobar el marco de referencia para las pruebas que se administren, establecer un plan de comunicación pública del sistema y resultados de las pruebas, supervisar los aspectos técnicos del proceso, etc.), garantizando de esta manera, que haya una buena sintonía entre el Ministerio y el Instituto.
- Junto con el Directorio que desarrolla funciones ejecutivas del Instituto, se debe considerar el establecimiento de un Consejo Consultivo constituido por miembros de destacada trayectoria y experiencia en la actividad educativa, nombrados por el Ministro de Educación, previa selección conforme al Sistema de Alta Dirección Pública. El Consejo Consultivo mantendrá un carácter plural y democrático.
  - Los planes de evaluación a mediano plazo deben ser aprobados por el Ministerio de Educación.
  - Las especificaciones y determinados instrumentos deben ser aprobados en coordinación con las unidades de currículo. Eso permite que dichas áreas se involucren naturalmente en la tarea de evaluación y participen del proceso.

---

<sup>6</sup> Al respecto se puede consultar el texto de Bracho (2022).

- El Instituto conformará Comités Técnicos sobre distintos aspectos de las Evaluaciones. Ello le permitirá fortalecer la legitimidad social de sus procesos frente al Ministerio y a la opinión pública.

Como vemos, el Ministerio de Educación mantiene funciones en la planificación de mediano plazo y en la articulación de los instrumentos de evaluación con el desarrollo del currículo.

La capacidad para crear y desarrollar experiencia técnica sólida y estable es uno de los principales retos que debe resolver el Instituto. Por esta razón, la autonomía administrativa y estabilidad presupuestaria debe orientarse en reclutar y mantener un número permanente y creciente de expertos. Para lo cual es necesario establecer procesos institucionales que permitan seleccionar los candidatos adecuados, entrenarlos y luego retenerlos dentro de la organización.

El perfil de los profesionales debe conjugar la solvencia de cuadros senior con amplia experiencia en sistemas de evaluación educativa junto con jóvenes profesionales de las disciplinas mencionadas interesados en aprender e investigar en medición de la calidad educativa. Es necesario identificar los incentivos, ya sean económicos, profesionales, laborales o académicos para fidelizar a los expertos del Instituto. Pues, la experiencia empírica señala que hay una fuerte tendencia a transitar hacia el sector privado u organismos internacionales cuando de obtener el conocimiento y la experiencia en medición educativa (Ferrer & Fiszbein, 2015).

Se debe considerar que la evaluación educativa es una tarea compleja y delicada, por lo tanto, requiere de un equipo multidisciplinario que incluya profesionales del campo de la psicología, educación, estadística, sociología, economía, comunicaciones, informática, administración, especialistas en gestión y mapeo de procesos, entre otros. A diferencia de años pasados, en que los especialistas adquirían sus conocimientos mientras trabajaban en las unidades de evaluación (por ejemplo, los que realizaban análisis psicométricos) ahora existen más profesionales que han recibido una formación específica en áreas relacionadas a la evaluación de aprendizajes. Así, el terreno para reclutar personal más capacitado se muestra más promisorio que el pasado.

En el caso de Perú, la Oficina de Medición de la Calidad de los Aprendizajes (UMC), que conduce el sistema de evaluación de logros de aprendizaje, reconocida como un área con altos niveles de eficiencia logró construir un equipo con alto nivel técnico debido a que sus tareas requerían habilidades complejas (Dargent 2014). En el momento en que empezó a funcionar la oficina no existía en las universidades peruanas profesionales con el conocimiento y expertise requeridos para el diseño e implementación de evaluaciones de logro de aprendizajes, ello obligó a que el fortalecimiento técnico se hiciera internamente a partir de un reclutamiento meritocrático de personal y de acceso a apoyos externos. Luego de una primera etapa en que la unidad recibió un acompañamiento cercano del Grupo de Análisis para el Desarrollo (GRADE) no tercerizó sus procesos lo que fortaleció a sus equipos técnicos. En esta fase los recursos financieros, especialmente los que provenían del Banco Mundial<sup>7</sup>, le permitió a la oficina, por un lado, contar con asesorías externas, y, otro lado, participar en evaluaciones internacionales (especialmente PISA), lo que nutrió y enriqueció sus procesos técnicos y logísticos. Además, sumó la cercanía

---

<sup>7</sup> El origen de la unidad estuvo ligado a un programa de mejoramiento de la educación básica el cual tenía como uno de sus componentes la evaluación de logro de aprendizajes.

con el grupo de trabajo sobre Estándares y Evaluación del Programa de Promoción de la Reforma Educativa para América Latina y el Caribe (PREAL) debido a que la coordinación estaba a cargo de GRADE, y de donde provinieron la mayoría de los expertos que prestaron asesoría a la unidad. También se puede señalar que un referente que se utilizó para evaluar las fortalezas y debilidades del sistema de evaluación fue el set de instrumentos y rúbricas del Systems Approach for Better Education Results (SABER).

## **Gobernanza: Propósitos**

Las características principales de los sistemas de evaluación estandarizada de logros de aprendizaje escolar están –o deberían ser– definidas en función de los propósitos y los mecanismos a través de los cuales se busca alcanzar dichos propósitos. Todos los sistemas de evaluación tienen como objetivo general, de una u otra forma, contribuir con la mejora de los sistemas educativos, en particular en lo que se refiere a los aprendizajes de los estudiantes. En lo que suelen diferenciarse los sistemas de evaluación es en la manera cómo buscan contribuir con la mejora educativa, en particular, los mecanismos que usan para ello.

En los años noventa cuando los países de la región instalaron sus sistemas de evaluación la reflexión sobre los fines específicos que se esperaba persiguiera dichos sistemas, y sobre las características técnicas que deberían tener los instrumentos para adecuarse a esos fines, fue casi inexistente. No obstante, luego de esa primera etapa de desarrollo lo que se ha observado es que la revisión de los objetivos es un proceso permanente en los sistemas de evaluación. En parte porque la propia práctica evaluativa interpela la definición de sus propósitos y estrategias definidas, pero también porque es común que “las autoridades ministeriales comiencen a demandar, sobre la marcha, que las evaluaciones sirvan para nuevos propósitos o que aporten información para fines para los que no fueron diseñadas” (Ravela et al., 2006, p. 58)

En ese marco Costa Rica no será la excepción a la situación descrita. De hecho, ya el sistema de macro evaluación de este país ha enfrentado desde su creación diversos dilemas que fueron apareciendo a lo largo de los años (Francis, 2023). Lo que queremos señalar es que la presión sobre las áreas de evaluación es una variable siempre presente. De allí la importancia de explicitar los objetivos que se buscan con los sistemas de evaluación. La información requerida para formular o retroalimentar políticas no es la misma que se necesita para devolver resultados a cada centro educativo para promover procesos de mejora pedagógica o institucional. Sin embargo, lo cierto es que la experiencia regional “muestra que hemos operado evaluaciones que no terminaron de ajustarse a los tiempos y enfoques demandados por las escuelas, y que, al mismo tiempo, fueron poco utilizadas en la formulación de políticas” (EIGE, 2010, p. 8). Ello parece indicar que el desafío mayor de los sistemas de evaluación hacia adelante es encontrar la manera de vincularse con los sentidos, la cultura y las prácticas de los actores educativos de distintos niveles.

En ese marco, se propone que Costa Rica opte por un sistema de evaluación de carácter comprensivo que incluya un conjunto diverso de propósitos y enfoques evaluativos, comprendiendo a los sumativos como a los formativos con el objetivo de brindar información pertinente que permita tomar decisiones en los distintos niveles para contribuir con el logro de

los objetivos generales de la educación. En otras palabras, se recomienda apostar por un diseño de evaluación balanceado que genere una combinación virtuosa entre evaluaciones sumativas de logro (al final del año escolar) con evaluaciones formativas gestionados por los propios equipos profesionales de los centros educativos (durante el año escolar). Esta recomendación cobra mayor relevancia en el caso que Costa Rica opte por la vía de constituir un organismo autónomo que se encargue de las evaluaciones dado que, como se ha mencionado anteriormente, un riesgo asociado a esa decisión es la debilidad en la articulación con los ministerios de educación.

Las razones que justifican esta propuesta son:

- Los balances realizados por los organismos internacionales (OECD, 2013; UNESCO, 2019) sobre diversos sistemas de evaluación coinciden en señalar que aún las evaluaciones que se conciben como de bajas consecuencias conllevan riesgos en sobre responsabilizar a los actores escolares por los resultados de las evaluaciones estandarizadas, que siempre son limitadas. En ese sentido, recomiendan que los países encuentren un balance entre evaluaciones sumativas y formativas en la revisión de sus sistemas de evaluación.
- La propuesta al incorporar el dispositivo de evaluaciones formativas (en formatos variados) es que esta actúe como la bisagra que articule la enseñanza con los aprendizajes. En países que tienen realidades de aprendizaje muy heterogéneas, como es el caso de Costa Rica, las evaluaciones sumativas permiten identificar la magnitud de los grupos de estudiantes que presentan dificultades de ciertos grados y competencias, “pero no posibilitan a los maestros conocer el estado por estudiante para tomar medidas pedagógicas [oportunas] a fin de movilizar los aprendizajes de manera formativa; por eso [estos instrumentos] tienen un gran potencial para ingresar profesional y formativamente en la conversación y la acción” (Henríquez, 2022, p. 258).
- Introducir las evaluaciones formativas en el sistema de evaluación permitirá, por un lado, atenuar el grado de responsabilización que traen consigo las evaluaciones estandarizadas en directivos y docentes de los centros educativos, y, por otro lado, aumentar la probabilidad que los informes de resultados sean mejor aprovechados por los actores escolares en tanto “la experiencia ha mostrado las dificultades para devolver la información a las escuelas, tanto en cuento al tiempo como a la pertinencia en el modo en que la misma ha sido comunicada” (EIGE, 2010, p. 15) De esta manera se puede reducir la distancia que actualmente existe entre las evaluaciones de sistema y las evaluaciones de aula (empezando por el lenguaje que cada una usa) que los docentes llevan a cabo cotidianamente (ver gráfico 1), y que constituyen uno de los mayores desafíos que enfrentan las evaluaciones de sistema, y que en algunos casos ha derivado en pérdidas de relevancia y legitimidad de estas<sup>8</sup>. Si bien estas dos evaluaciones tienen propósitos y características distintas, consideramos que la evaluación aplicada a la enseñanza y el aprendizaje consiste en un proceso sistemático y riguroso de obtención de datos, de manera que sea posible disponer de información continua y significativa para diagnosticar, formar juicios de valor con respecto a ella y tomar las decisiones adecuadas para mejorarla, la información obtenida a través de procedimientos e

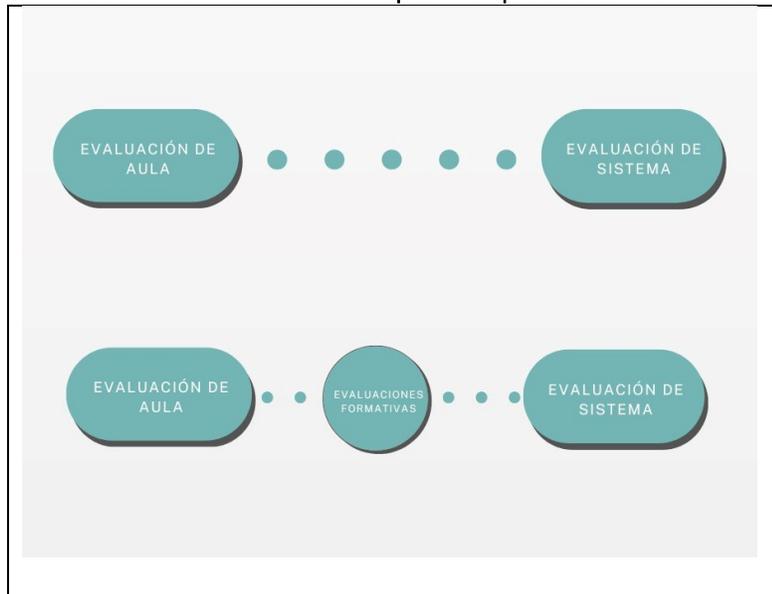
---

<sup>8</sup> Un ejemplo de esto fue la campaña Alto al SIMCE que surgió en 2013 promovida por organizaciones de docentes, estudiantes y familias que rechazaban los efectos del sistema de evaluación en la educación chilena.

instrumentos diferenciados entre evaluaciones de sistema y evaluaciones de aula debería ser coincidente en la medida que son dos miradas a un mismo proceso.

Gráfico 2

La evaluación formativa como dispositivo que acerca la evaluación de sistema y de aula



Fuente: Elaboración propia.

## Las evaluaciones sumativas

Costa Rica tiene un recorrido en el ámbito evaluativo, por lo que no es que parta de cero. Sin embargo, en la revisión y definición que ha emprendido sobre su sistema de evaluación es clave hacer algunas definiciones como (i) los grados y áreas curriculares que serán objeto de evaluación; (ii) la periodicidad de la evaluación; (iv) la cobertura poblacional; (v) el nivel de resolución de los resultados (implicancias técnicas); (vi) las principales desagregaciones y modos de informar los resultados; (vii) el nivel de publicación de los resultados; y (viii) la participación del país en mediciones internacionales de aprendizaje. En el Anexo 2. se presentan recomendaciones relacionados a algunos de estos aspectos.

## Evaluaciones formativas

Algunos países como Chile, Perú y Uruguay ofrecen algún tipo de herramienta de carácter formativo. Estas consisten en la entrega de recursos de evaluación para que los docentes puedan aplicarlos y analizarlos de manera flexible con sus estudiantes con el propósito de retroalimentar las prácticas pedagógicas. Es importante mencionar que aun cuando estos instrumentos fueron diseñados antes de la pandemia por el COVID 19 su uso durante el periodo de educación remota se potenció, y, pueden constituir herramientas útiles para situaciones de emergencia (desastres naturales, huelgas prolongadas, etc.)

En el caso de Chile, la Agencia de Calidad de la Educación ofrece el Diagnóstico Integral de Aprendizajes (DIA)<sup>9</sup> que es una herramienta voluntaria de uso formativo para que los docentes puedan monitorear el aprendizaje de sus estudiantes en tres momentos a lo largo del año escolar: al inicio, en el intermedio y al cierre. La herramienta consta de (i) pruebas y cuestionarios (están disponibles tanto en versión impresa como en línea), (ii) reportes de resultados (a nivel de sección y centro educativo) y (iii) orientaciones pedagógicas para el análisis y uso de resultados. El DIA “pretende activar ciclos de reflexión pedagógica en las escuelas durante las tres instancias de aplicación en el año, que conducen a la planificación de acciones para la mejora de los aprendizajes. Esta reflexión debería realizarse de manera colaborativa, donde los docentes sientan el apoyo de equipos directivos, padres o apoderados y sus pares, para tener la confianza de lograr las metas” (Henríquez, 2022, p. 254)

Por su lado Uruguay ha desarrollado desde 2014 el Sistema de Evaluación de Aprendizaje (SEA)<sup>10</sup>. Esta herramienta ofrece pruebas entendidas como evaluaciones *para* el aprendizaje (a diferencia de las evaluaciones de sistema que serían evaluaciones *del* aprendizaje). Lo distintivo del caso uruguayo es que las pruebas “se aplican en línea aprovechando la infraestructura tecnológica instalada en los centros educativos por el Plan Ceibal, para reflexionar e intervenir sobre los aprendizajes de los estudiantes a escala nacional. El formato en línea permite tener mayor cobertura, resultados al instante, y un gran ahorro en papel”.

La aplicación en línea del SEA permite retroalimentar con información en tiempo real al estudiante sobre su aprendizaje, al docente sobre su grupo de estudiantes, al director sobre su centro educativo y a los inspectores a nivel de su jurisdicción. De allí que el objetivo de estos recursos formativos esté conceptualizado como la generación de discusiones colectivas y reflexiones sobre la enseñanza en cada centro educativo. En el caso específico de los docentes se espera que, en tantos profesionales, estos sumen información que les brinda el SEA a otros insumos como los trabajos, tareas y evaluaciones en aula para que tomen decisiones. El trabajo sobre patrones de error y el uso de evidencia para la planificación son algunos de los efectos que se han observado en el uso del SEA (Peri, 2016). De otro lado, en los últimos años, SEA ha desarrollado otros componentes innovadores que están en consonancia con las tendencias futuras de los sistemas de evaluación. Así, el sistema brinda la posibilidad de que el docente pueda seleccionar actividades para diseñar sus propias pruebas de tal manera que sean pertinentes para el grupo de estudiantes que atiende. Asimismo, ofrece evaluaciones adaptativas (SEA+) lo que permite una mayor precisión en informar sobre el desempeño particular de cada estudiante al liberarse del efecto techo o piso que tiende a caracterizar a las evaluaciones estandarizadas de lápiz y papel.

En el caso de Perú<sup>11</sup>, los recursos de evaluación formativa no han sido incorporados de manera explícita como un componente de las evaluaciones de sistema, sin embargo, su concepción y elaboración ha estado a cargo de la Oficina de Medición de la Calidad de los Aprendizajes (UMC), instancia encargada de las mismas. El Kit de evaluación diagnóstica ofrece a los docentes instrumentos que les permita conocer y monitorear el nivel de aprendizaje de sus estudiantes.

---

<sup>9</sup> La información sobre el DIA se puede consultar en: <https://diagnosticointegral.agenciaeducacion.cl/>

<sup>10</sup> La información sobre el SEA se puede consultar en: <https://sea.anep.edu.uy/>

<sup>11</sup> La información sobre estos instrumentos se puede consultar en: <https://repositorio.perueduca.pe/docentes/planificacion/primaria-herramientas.html>

Consta de tres herramientas impresas: (i) pruebas, (ii) manual de uso de las pruebas y (iii) registro de respuestas para el docente. El Ministerio de Educación envía estos instrumentos a todas las escuelas del país, y son los docentes los que se encargan de su aplicación y análisis. En las primeras versiones de esta herramienta se potenció su aplicación a través de la celebración del Día de Logro que era uno de los momentos de la Movilización Nacional por la Mejora de los Aprendizajes en la que los estudiantes mostraban sus trabajos y resultados, y cuya implementación estaba a cargo de las instancias de gestión educativa regional y local.

En suma, se propone que Costa Rica incluya progresivamente en el diseño de su sistema de evaluación la entrega de dispositivos de evaluación formativa a los centros escolares. De las experiencias revisadas el modelo más consolidado y valioso por su potencial efecto positivo en la mejora de las prácticas pedagógicas es el SEA de Uruguay, sin embargo, es claro que su desarrollo tiene a la base un conjunto de condiciones difíciles de replicar, como la expansión de las TIC a través del Plan Ceibal, y la sólida tradición de evaluación de aprendizajes de carácter formativo que ha caracterizado al sistema educativo uruguayo. En ese sentido, un referente intermedio constituye el modelo chileno. La construcción de instrumentos puede iniciarse con la selección de algunos grados y competencias en la que los equipos técnicos de Costa Rica tengan más solvencia. Dado que este tipo de instrumentos no conllevan las exigencias técnicas (psicométricas) exigidas para las evaluaciones sumativas se pueden aprovechar para elaborar los instrumentos a través de un plan de formación de capacidades con docentes, formadores y supervisores. Al igual que en Uruguay se debe incluir la mirada disciplinar y didáctica de estos actores en la construcción de las actividades y reactivos de evaluación. Si bien en el siguiente acápite nos referiremos nos parece importante no dejar de señalar que es insoslayable la inclusión de las TIC en la definición del nuevo sistema de evaluación costarricense.

**Cuadro 1**

**Recomendaciones sobre componentes y características del sistema de evaluación de aprendizajes**

Aspecto	Evaluaciones de Sistema <sup>a/</sup>	
	Evaluaciones Sumativas	Evaluaciones Formativas
¿Cuál es el propósito?	Evaluar el logro de aprendizajes en diferentes competencias y grados del Currículo Nacional con el objetivo de contribuir a la toma de decisiones para la mejora de la calidad de la educación.	Evaluar el progreso del aprendizaje.
¿Cuándo se aplica?	Al finalizar el año escolar.	En diferentes momentos del año escolar (pueden ser dos o tres aplicaciones)
¿Quién lo diseña?	La entidad encargada de las evaluaciones.	La entidad encargada de las evaluaciones de sistema.
¿Quién lo aplica y analiza?		Los docentes de cada centro escolar.
¿Qué información entrega?	<ul style="list-style-type: none"> <li>• Resultados de logros para los distintos estratos definidos (según sea censal/muestral)</li> <li>• Comparación en el tiempo.</li> </ul>	<ul style="list-style-type: none"> <li>• Información a nivel de estudiante y de aula.</li> <li>• Progresos obtenidos por cada estudiante.</li> </ul>
¿Qué usos <sup>b/</sup> tiene?	<ul style="list-style-type: none"> <li>• Monitoreo de resultados del sistema y tendencias en el tiempo.</li> <li>• Retroalimentación de políticas: curricular, materiales educativos, formar y fortalecer la profesionalización docente, gestión escolar, etc.</li> <li>• Retroalimentación de prácticas pedagógicas y de gestión educativa.</li> <li>• Focalización de recursos y apoyos para centros.</li> <li>• Establecimiento de metas en diferentes niveles de gestión.</li> <li>• Identificación de buenas prácticas.</li> <li>• Movilización social por los aprendizajes.</li> <li>• Información para diseño, seguimiento y evaluación de intervenciones.</li> </ul>	<ul style="list-style-type: none"> <li>• Identificar grupo de estudiantes que requieren apoyos adicionales.</li> <li>• Reflexión y mejora de estrategias pedagógicas y didácticas utilizadas.</li> </ul>

a/Las evaluaciones internacionales que también forman parte de las evaluaciones de sistema se desarrollan en el siguiente acápite.

b/ Los usos varían según la cobertura de la evaluación sea censal o muestral.

Fuente: Elaboración propia.

Hemos dejado para el final la recomendación sobre la inclusión de la evaluación de certificación dado que es un tema complejo. Resulta difícil hacer una recomendación concluyente sin tomar en cuenta las tradiciones y concepciones que tiene la sociedad costarricense sobre las evaluaciones sumativas que parecieran tener arraigo (Francis, 2023). De otro lado, como señala Pinkasz (2021) actualmente es difícil hablar de modelos de evaluación binarios que clasifiquen a los sistemas como de altas (sumativas) o bajas consecuencias (formativas). En la realidad lo que se observa es que en los sistemas hay una coexistencia, predominio o tensiones entre ambas funciones, “algunas funciones tienen consecuencias para los individuos, pero no necesariamente caracterizan al sistema, como es el uso de algunas pruebas para la certificación de estudios que están orientadas a la evaluación de las personas y no a la de los sistemas” (p. 34).

En este marco, y consistente con la recomendación hecha de que Costa Rica apueste por un modelo de evaluación comprensivo resulta más útil pensar en qué variables pueden moderar o amortiguar los posibles efectos negativos que podrían traer una prueba de certificación.

Los factores amortiguadores son:

- Se debe utilizar en los ciclos superiores del sistema educativo. Por lo tanto, no es aconsejable que se aplique al final de la educación general básica.
- Los estudiantes deben tener más de una oportunidad de rendir las pruebas y se les debe considerar el mejor puntaje obtenido.
- Los resultados de la evaluación deben constituir solo una parte del proceso de certificación de aprendizajes de los estudiantes.
- La evaluación debería ser flexible. Se puede pensar en tener un módulo común y otro variable.
- Respecto al alineamiento curricular de las pruebas existe dos posiciones. La primera enfocada en el estudiante señala que es mejor evaluar habilidades generales (llamadas a veces aptitudes) dado que no todos los centros cubren los contenidos curriculares exigidos lo que genera una desventaja en los estudiantes que no han estado expuestos a dichos contenidos. La segunda posición, centrada en el sistema, enfatiza que las pruebas deben tener una referencia curricular dado que las pruebas lo que hacen es dar una señal de lo que se debe enseñar y lo que importa aprender. Genera un incentivo a para implementar el currículo en todo el sistema, y por lo tanto es mejor apostar por esta posición. Desde un punto de vista técnico, como todo instrumento de medición, y en este caso más, debe contar con investigación regular que entregue evidencia acerca de la validez de las pruebas que se construyan. Por otra parte, dado que las pruebas se emplearán con fines de selección, es necesario aportar evidencia acerca de la capacidad predictiva de las pruebas, analizando la correlación entre las pruebas y otros criterios de selección (como las notas de la enseñanza secundaria) con respecto al rendimiento académico en la universidad (notas y otros criterios como las tasas de graduación).

Perú no cuenta con pruebas de certificación. En algún momento se evaluó tener una prueba de certificación al final de la educación básica con el propósito de que sus resultados sirvan para la admisión a la educación superior. Sin embargo, no se llegó a aplicar para este objetivo.

En una consultoría realizada por la Universidad del Pacífico & el Centro de Medición de la Universidad Católica de Chile (2017) se hicieron las siguientes recomendaciones para contar con una prueba estandarizada para el ingreso a la educación superior, varias de las cuales coinciden con las señaladas como factores amortiguadores: (i) las áreas que se evalúen en un comienzo debieran ser: lectura, matemática y formación ciudadana, con miras a la futura inclusión de áreas adicionales tales como ciencias y producción escrita; (ii) los resultados de las pruebas constituirán solo uno de los insumos que conformarían un sistema de selección para la educación superior; (iii) las pruebas deben tener una referencia curricular de manera que se busque evaluar el manejo de los postulantes respecto a contenidos y habilidades definidos en el currículo nacional. Considerando que el currículo peruano tiene un enfoque por competencias ello permite articularla con la educación superior y aumenta las probabilidades de evaluar aspectos sobre los cuales los estudiantes han tenido oportunidades de aprender.

## **Técnicas e instrumentos de medición**

Una vez definido los propósitos a los que debe responder un sistema de evaluación resulta necesario que éstos se traduzcan en una selección y construcción de instrumentos y formatos que cuenten con estándares adecuados de calidad técnica que incluya consideraciones para garantizar una evaluación justa y equitativa. Sobre este tema, en esta sección de manera puntual se presentarán sugerencias respecto al marco de referencia de la evaluación, el uso de TIC, y la inclusión de instrumentos para contextualizar los resultados educativos.

## **Referentes conceptuales para la evaluación: alineamiento curricular y progresiones de aprendizaje**

Las evaluaciones estandarizadas son útiles en tanto brindan información para el diseño e implementación de la política educativa y para el cambio de las prácticas pedagógicas en los centros educativos. En ese sentido, una apuesta por la evaluación que se aproxime a la medición de los aprendizajes que se quiere promover como sociedad y que suelen estar incluidos en el currículo es la que brinda insumos relevantes para la mejora educativa. En ese marco, se propone que el sistema de macro-evaluación de Costa Rica tenga como referencia los objetivos curriculares. Esta apuesta se puede complementar con la participación en evaluaciones internacionales, como PISA<sup>12</sup>, que no responden a un referente curricular en tanto es una evaluación de los conocimientos y destrezas que un estudiante próximo a concluir su

---

<sup>12</sup> Otra evaluación que Costa Rica puede considerar en esta perspectiva es el Estudio Internacional de Alfabetización Computacional y Manejo de Información, ICILS por sus siglas en inglés (International Computer and Information Literacy Study). Es un estudio internacional comparativo basado en computadora que mide cuán preparados están los estudiantes en el uso de computadoras para manejar información. Este estudio es liderado y desarrollado por la Asociación Internacional para la Evaluación del Logro Educativo (IEA).

escolaridad obligatoria debería haber alcanzado considerando los desafíos que debe enfrentar en el futuro.

Las razones principales que justifican dicha propuesta son: (i) una evaluación de calidad requiere fundamentarse en expectativas claras y explícitas sobre los aprendizajes de los estudiantes para garantizar la validez, confiabilidad, transparencia y usabilidad de las evaluaciones (OECD, 2013); (ii) las evaluaciones constituyen un medio para precisar a los docentes y familias qué es lo que se espera que los estudiantes alcancen en cada competencia y ciclo escolar, y de esa manera genera también un incentivo para la concreción del currículo en aula y contribuye la demanda ciudadana; (iii) promueve el alineamiento de los diferentes componentes de la política pedagógica (evaluación de aprendizajes, la formación docente y la elaboración de materiales y recursos educativos) de tal manera que los docentes reciban mensajes consistentes y aumente la probabilidad de que usen los resultados de las evaluaciones.

Ahora bien, no siempre las expectativas de aprendizajes están claramente definidas en los currículos nacionales. En algunos países, como en el Perú, el Currículo Nacional de Educación Básica (CNEB), además de identificar las principales competencias que deben desarrollar los estudiantes durante su escolaridad obligatoria incluye estándares de aprendizaje a través de descripciones del progreso típico de cada competencia en niveles de creciente complejidad desde el inicio hasta el fin de la escolaridad (ver Anexo 3).

El proceso de elaboración de los estándares de aprendizaje en Perú empezó el 2009 y fue un proceso complejo<sup>13</sup> debido a los desafíos técnicos que supone la construcción de este tipo de instrumentos y su relación con el Currículo Nacional, pero también a los retos organizacionales y políticos relacionados con la coordinación y acople del rol del Ministerio de Educación en este proceso.

Una de las decisiones más importantes que se tuvo que tomar para la elaboración de los estándares estuvo relacionada con la elección del enfoque que adoptarían los estándares y que debía responder al uso que se haría de ellos. Luego de revisar diversos modelos de estándares se optó por el modelo australiano de mapas de progreso por sus beneficios pedagógicos y porque era sensible a la diversidad de niveles de logro que se evidencian en las aulas peruanas (Tapia et al, 2016). Este modelo describe la forma en que los estudiantes típicamente se mueven a través del aprendizaje en las diferentes competencias curriculares. Entre otras cosas permite contar con una hoja de ruta para identificar el conjunto de habilidades y conocimientos que los estudiantes deben dominar para volverse competentes en resultados curriculares más complejos.

En caso Costa Rica decida optar por la construcción de estándares de aprendizaje se pone a consideración las siguientes recomendaciones que fueron las que los expertos internacionales recomendaron al Perú: (i) revisar diversos modelos o propuestas de estándares; (ii) conformar un equipo técnico elaborador de los estándares con perfiles variados (especialistas en currículo, en evaluación y expertos de áreas académicas); (iii) contar con un equipo de asesores y externos nacionales e internacionales; (iii) planificar desde el inicio una campaña de comunicación dirigida al sector educativo y opinión pública con el fin de que comprendan que la importancia

---

<sup>13</sup> Un detallado recuento de este proceso se encuentra en Tapia et al (2016).

de la elaboración de estándares descansa en un objetivo fundamental: mejorar la calidad educativa de los estudiantes; (iv) Desarrollar estrategias que permitan enfrentar un proceso técnico largo y laborioso que demanda precisiones y decisiones críticas que deben adoptarse desde el principio y continuamente, (v) Diseñar e implementar procesos de consulta y de concertación con los actores sociales, pues ello resulta fundamental para la aceptación y legitimación social de los estándares (Tapia et al., 2016).

### **TIC: Hacia formas más innovadoras de evaluar**

Actualmente los sistemas de evaluación usan de manera predominantemente instrumentos de lápiz y papel. Sin embargo, crecientemente algunos programas internacionales de evaluación, como el caso de PISA, están optando por incluir formatos innovadores en base al uso de TIC, o evaluaciones en línea, como es el caso de Uruguay. Estos nuevos entornos y formatos ofrecen un gran potencial para aproximarse de manera más pertinente a los objetivos de aprendizaje que se valoran en el sistema educativo, y marcan el derrotero hacia donde se deben orientar los sistemas de evaluación. En ese sentido, Costa Rica, en el rediseño de su macro evaluación debe evaluar cómo empezar a incluir las tecnologías de información y comunicación tanto en sus formatos de evaluación<sup>14</sup>, pero también en sus procesos de aplicación de las pruebas, captura de los resultados y procesamiento y análisis en tanto supone una reducción notable de los costos, además de facilitar el intercambio y gestión de una gran cantidad de datos. Este debería ser un aspecto no negociable en tanto está directamente relacionado con la utilidad y sostenibilidad del sistema de evaluación.

Desde la perspectiva del diseño de instrumentos, entre las ventajas de la inclusión de las TIC en los sistemas de evaluación se pueden citar: (i) el diseño de pruebas de mejor calidad en tanto permite incluir estímulos, simulaciones, entornos interactivos que se acercan más a los desafíos reales que los estudiantes enfrentarán en su vida laboral y ciudadana, y a la vez permite evaluar niveles cognitivos complejos y ampliar las competencias o aprendizajes que muchas veces no son posibles de evaluar usando formatos de lápiz y papel; (ii) el uso de test adaptativos informatizados (TEI) que si cuentan con un amplio rango de dificultad de ítemes pueden ser más efectivos para medir el nivel de aprendizaje de los estudiantes en tanto es menos probable que presenten efectos “techo” o “piso” (Burga, 2019), lo que favorece a las poblaciones de muy bajo o muy alto rendimiento; (iii) la tecnología permite que los formatos se adapten a las necesidades diversas de los estudiantes, por ejemplo, el uso de tablets que agrandan la letra para personas con discapacidad visual. Por su potencial de adaptación, facilidad de uso y eliminación de barreras las TIC garantizan una evaluación más justa para todos los estudiantes.

---

<sup>14</sup> Una posibilidad es empezar con su inclusión en los dispositivos de evaluación formativa como una suerte de ir probando para que luego que se madure y consolide el proceso pasar a los diseños de evaluación con consecuencias.

Cabe señalar que en Perú no se ha incluido evaluaciones por computadora, salvo para la implementación de PISA y para realizar el Estudio Virtual de Aprendizajes (EVA)<sup>15</sup> durante la educación remota por el COVID 19. Esta última no pudo tener representatividad de la población escolar debido a las limitaciones de acceso a tecnologías de las familias y escuelas.

### **Factores asociados: presentar resultados comparables**

El término de factores asociados alude a la aplicación de instrumentos, generalmente cuestionarios, a través de los cuales se recoge información sobre las características individuales y familiares de los estudiantes, así como de las características de las escuelas: sus actores y procesos e incluso de los entornos en los que se desenvuelven las instituciones escolares. El propósito principal es utilizar esta información para *investigar* su asociación con los resultados medidos por las pruebas estandarizadas. Sobre las dimensiones y aspectos que se deben considerar existe más o menos un consenso en la literatura. Los desafíos sobre este tema van más en la dirección de cómo aproximarse con instrumentos “masivos” a dimensiones relacionadas al *proceso* de enseñanza aprendizaje que tienen, según los marcos conceptuales, un rol clave en el logro de resultados educativos, pero, que, por la complejidad que demanda su operacionalización y recojo generalmente no son incluidos como variables de estudio. El desempeño docente es un ejemplo de este tipo de constructos que tienen, además, el potencial de retroalimentar las prácticas escolares y la política educativa.

Abordar este reto, supone, entre otras cosas, planificar bien los estudios de factores asociados. En ese sentido, se recomienda incluir el programa de investigación de factores asociados en el plan de evaluaciones a mediano plazo. Los estudios de factores asociados deben tener sus propios protocolos técnicos al igual que las pruebas: marcos de fundamentación, objetivos, matriz de especificaciones, definiciones, protocolos de aplicación, plan de análisis, etc. Muchos sistemas de evaluación tienden a descuidar este aspecto.

Si se tienen evaluaciones anuales habría que decidir cuál es la periodicidad más adecuada que deberían tener los estudios de factores. ¿Tiene sentido realizar estudios de factores todos los años? Los estudios de factores asociados responden a una lógica correlacional, es decir, responden sobre el qué. En ese sentido, se sugiere enfocarse en pocas dimensiones, esto ayuda también al momento de comunicar los resultados. En cada iteración o ciclo evaluativo se puede ir variando de foco de análisis y profundizando dicha dimensión.

De otro lado, no toda la información se tiene que recoger ad hoc junto con la aplicación de los instrumentos de evaluación es importante también hacer uso de otras bases de datos disponibles con las cuales se puede relacionar la información del rendimiento. Los sistemas de estadística educativa, las evaluaciones docentes y directivas y las encuestas de hogares incluyen cada vez más información relevante para la política educativa.

De otro lado, los sistemas de evaluación deben preocuparse por recoger información sobre variables que sitúen o analicen las condiciones o circunstancias que permitan comprender mejor los resultados escolares. Como señala uno de los estándares del INEE (2014), una evaluación

---

<sup>15</sup> <http://umc.minedu.gob.pe/resultadoseva2021/>

estandarizada rigurosa requiere contar con información para contextualizar los resultados de los estudiantes:

“Ninguna decisión que tenga consecuencias importantes sobre los individuos o instituciones, se basará únicamente en los resultados de solo un instrumento de evaluación, por lo cual, deberán considerarse otras fuentes confiables de información que incrementen la validez de las decisiones que se tomen...” (p. 15)

En ese marco los sistemas de evaluación deben difundir sus resultados incluyendo algunas variables de control. En los países de América Latina el principal determinante de los resultados es la composición social de los estudiantes, por lo tanto, toda comparación de resultados entre escuelas, regiones o provincias debe realizarse controlando por dichas características (Ravela et al, 2008). En ese sentido, Costa Rica debería tomar en cuenta esta sugerencia en el diseño de su sistema de evaluación. Ello permite, por ejemplo, que la comparación del desempeño entre las escuelas privadas y públicas sea más justa. Además, contribuye a centrar la discusión en lo que realmente puede hacer el sistema educativo para mejorar los resultados descontando el efecto del origen social de los estudiantes a la vez que amplía y desplaza la exigencia ciudadana a las políticas económicas y sociales.

## **Difusión y uso de resultados**

La evaluación por sí sola no mejora los aprendizajes. Su iteración tampoco. Como se ha mencionado en la primera sección del presente documento todos los sistemas de evaluación, de una u otra forma, tienen como objetivo general contribuir con información para la toma de decisiones de mejora de la calidad de los sistemas educativos. No obstante, la experiencia por la que han transitado los sistemas de evaluación con cierta trayectoria evidencia que existe una brecha entre la intencionalidad declarada por los sistemas y el aprovechamiento efectivo que hacen los destinatarios de la información que brindan los operativos de evaluación (Pinkasz, 2021; Sempé y Andrade, 2017). Ciertamente, el uso de la información que hacen los países varía según los propósitos específicos definidos en sus modelos evaluativos sobre la atribución a las consecuencias de los resultados de las evaluaciones. Además, de otras características de los propios países, como, por ejemplo, su carácter federal, o factores institucionales técnicos, políticos e incluso comunicacionales de los propios sistemas de evaluación (OCDE, 2013; Pinkasz, 2021).

De otro lado, como lo menciona claramente el informe de EIGE (2011), en los últimos años se observa:

Este cambio en el foco de la devolución de resultados de los equipos ministeriales centrales y regionales para retroalimentar las iniciativas de política educativa hacia los directores, docentes (e incluso familias) para la mejora pedagógica e institucional de las escuelas complejiza la labor de las unidades de evaluación. Pone en tensión no solo la dinámica del proceso de difusión y uso de resultados<sup>16</sup> sino también el rol de las propias instancias evaluativas al interior de la estructura organizativa de los sistemas educativos (EIGE, 2011).

En este marco en que las demandas a las áreas de evaluación se reorientan a incidir en el cambio de los espacios escolares, el diseño de la macroevaluación de Costa Rica debe considerar esta variable y la ruta más efectiva en esa dirección es la de apostar al uso formativo de los insumos que provee la evaluación estandarizada. En ese sentido, debe partir de la premisa (realista) que la teoría de cambio que vincula el uso de los resultados de las evaluaciones estandarizadas con mejora educativa no es inmediata ni directa. Es una relación que depende de que se den ciertas condiciones, capacidades y actitudes en los actores involucrados, que a su vez están influenciados por sus propias características, concepciones e intereses, y que intervienen en una cadena compleja de procesos y efectos esperados o supuestos. En efecto, no se debe perder de vista que el sistema educativo tiene una estructura *anidada*, es decir los actores y organizaciones se encuentran contenidas unos dentro de otros e interrelacionados, y las interacciones entre los distintos niveles no siempre van en la misma dirección.

A continuación, se proponen algunas recomendaciones tomando la experiencia por la que ha transitado Perú desde el 2007 en la difusión y uso de resultados de las evaluaciones de rendimiento, en especial de la Evaluación Censal de Estudiantes (ECE). Por razones de espacio nos limitaremos al uso de resultados para incidir en la práctica pedagógica y en políticas educativas.

### **Uso de resultados para mejorar la práctica pedagógica**

La UMC elabora y distribuye anualmente informes de resultados de la ECE para las instancias regionales y locales de gestión educativa descentralizada, así como para todos los centros educativos evaluados. En este último caso reparte cuatro tipos de informes dirigidos: (i) al Director; (ii) a los Docentes de los grados evaluados y de los grados siguientes a la evaluación (informes pedagógicos por cada una de las áreas evaluadas); (iii) a los Padres de Familia, y (iv) una guía para desarrollar la Jornada de Reflexión sobre los resultados.<sup>17</sup> Cabe destacar que los informes para docentes que ha desarrollado Perú contienen sugerencias pedagógicas y propuestas de actividades ejemplificadas y modeladas que no está presente en ningún otro recurso curricular que llegue a las aulas.

La hipótesis a la base de esta práctica era que luego de la entrega de los informes ocurra una serie de cambios en el ámbito de la gestión del centro educativo, en el clima escolar y especialmente en el ámbito pedagógico, que se debía expresar en nuevas prácticas y acciones de docentes, directivos y familias. En conjunto, se esperaba que la información brindada genere

---

<sup>16</sup> Es importante puntualizar que esto no ocurre en las evaluaciones de altas consecuencias en los individuos, pues el resultado en estos casos tiene una relación directa con el uso, por ejemplo, la certificación de un estudiante en función del puntaje obtenido en un examen de final de nivel.

<sup>17</sup> Un ejemplo del set de informes de resultados de la ECE se puede encontrar en: <http://umc.minedu.gob.pe/resultadosece2016/>

una dinámica de cambio en la institución y en la mejora de las oportunidades de aprendizaje de los estudiantes. Sin embargo, el envío iterativo de informes escritos anuales a las escuelas no ha resultado ser suficiente para promover su uso e impulsar el cambio significativo de prácticas pedagógicas e institucionales por parte de los actores escolares (Sempé y Andrade, 2017)<sup>18</sup>.

En el mejor de los casos los informes son leídos y los resultados son compartidos con la comunidad escolar, e incluso se han generado ciertas acciones y cambios interesantes en algunos centros. No obstante, para hacer un salto cualitativo en este proceso hace falta complementar el esfuerzo de la entrega de informes con la generación de espacios para compartir y trabajar directamente con los equipos escolares utilizando los propios instrumentos y/o los resultados como una suerte de *pretexto* para catalizar la reflexión y el cambio de las prácticas de aula, y en consecuencia apuntar a la mejora de los aprendizajes.

Desde esa premisa se proponen tres ejemplos que se sitúan en la perspectiva de *formación de usuarios* donde la información que brindan las evaluaciones estandarizadas (entendida de manera amplia, no solo de resultados) son utilizadas formativamente para identificar necesidades de aprendizaje y ajustar las estrategias de enseñanza. Estas propuestas no han sido aplicadas a gran escala, pero consideramos que tienen un gran potencial formativo y podrían engarzar con las concepciones e intereses de los docentes e impulsar un cambio cualitativo de las prácticas pedagógicas.

Antes de presentar cada una de las propuestas, debemos señalar que otros países han optado por elaborar guías y otros recursos para ayudar a los actores escolares en la lectura e interpretación de los resultados de las evaluaciones estandarizadas<sup>19</sup>. Si bien estos recursos son muy útiles, desde la experiencia peruana, consideramos que habría que evaluar el costo beneficio de incluir más material escrito que competirá con la lectura y atención de los propios informes de resultados, además de que estos recursos no necesariamente empatan con los intereses más inmediatos de directivos y docentes. La experiencia nos indica que la mediación de un profesional es el factor catalizador para el uso provechoso de los resultados de las evaluaciones. En todo caso, dichos recursos pueden ser usados como un insumo en los espacios de mediación. Ahora bien, somos conscientes que la dificultad mayor para la sostenibilidad de estas propuestas radica en persuadir a las áreas de formación docente para que las incluyan en sus programas formativos.

---

<sup>18</sup> Existen varios factores que explican este hecho, entre ellos, quizá el más importante haya sido la debilidad de la política de formación docente. Tener un impacto en el cambio de prácticas pedagógicas supone al menos una sintonía de la política curricular, docente y de gestión, es decir un trabajo del conjunto de los distintos actores del sistema, no se puede hacer exclusivamente desde las áreas de evaluación.

<sup>19</sup> Un ejemplo de ellos son las guías que ha elaborado la Agencia de Calidad de la Educación de Chile, que se pueden consultar en el siguiente enlace: <https://bit.ly/427RweG>

- (i) Entrega de dispositivos de evaluaciones formativas.  
La utilidad de esta herramienta se ha detallado en la sección 3. Solo agregaremos que existe evidencia robusta que indica que incluir pruebas como un recurso pedagógico en el aula es una forma eficiente para aprender (Yang et al., 2021)
  
- (ii) Talleres de análisis de dificultad de ítems.  
A partir de las evidencias que ofrecen las evaluaciones, sobre aquello que pueden realizar los estudiantes y aquello que les cuesta más, por ejemplo, en lectura, los docentes pueden reflexionar acerca de qué aspectos deben trabajar para llevar a todos los estudiantes hacia el logro de los aprendizajes esperados. La dinámica del análisis de dificultad de los ítems *obliga* a que los docentes miren desde la lógica del aprendizaje de los estudiantes y no desde la enseñanza. Los orienta para redireccionar la pedagogía y tomar decisiones informadas, a partir de evidencia. En ese sentido, esta actividad cuestiona ciertos mitos asentados, como, por ejemplo, que las tareas de obtención de información siempre son las más sencillas cuando en realidad estas preguntas pueden complejizarse según las características del texto o el tipo de conexiones que hace el estudiante para ubicar la información solicitada, cuán explícita es la información que se busca o cuán evidente o entretejida está en el texto.

## Cuadro 2

### Talleres de análisis de dificultad de ítems

Objetivo de la actividad	(i) Analizar los ítems de la prueba en términos de los aprendizajes implicados, por ejemplo, de la competencia lectora, lo que específicamente pudieron realizar los estudiantes y las principales dificultades que tuvieron. (ii) A partir de ese análisis, reflexionar sobre cómo la enseñanza puede modificarse para atender mejor a los estudiantes.
¿En qué consiste?	Es una adaptación del modelo de taller de puntos de corte utilizando metodología Bookmark <sup>20</sup> . <ul style="list-style-type: none"><li>- En el grupo la moderadora (que debe ser una especialista de la prueba que se usará en la actividad) explica a los participantes el objetivo de la actividad y cuál será la dinámica que se seguirá.</li><li>- Se distribuye a los docentes un cuadernillo de ítems de lectura [que fue aplicado] ordenado por dificultad (estando esta dificultad estimada mediante procedimientos de teoría de repuesta al ítem), desde el ítem que resultó más fácil hasta el ítem que resultó más difícil.</li><li>- Se solicita que los docentes resuelvan los ítems de la prueba.</li><li>- Una vez que han terminado con la resolución de la prueba, la moderadora les pide a los docentes que analicen individualmente el cuadernillo de ítems ordenados de la prueba. Empiezan por el ítem 2 y responden a la pregunta: ¿Qué hizo que este ítem sea más difícil que el anterior? Y así se avanza sucesivamente con el ítem 3, 4, etc. Se les pide que escriban brevemente las razones señaladas debajo de cada ítem. Un aspecto clave para esta actividad es que la moderadora oriente/” obligue” a los docentes a que las razones que brinden sean pedagógicas y relacionadas con los aprendizajes de lectura, que se focalicen en describir los aprendizajes implicados en cada ítem y en explicar en qué medida estos pueden haber representado una dificultad para los estudiantes. Esto porque suele ser común que los docentes tiendan a explicar la dificultad de cada ítem aludiendo solo a que “los estudiantes no tuvieron oportunidad de aprender esos aprendizajes”, “los estudiantes tienen muchas carencias en sus hogares y por eso no pueden aprender bien”. Estos argumentos aluden a condiciones generales del aprendizaje, que son sin duda importantes, pero en este ejercicio se deben dejar fuera.</li><li>- Luego en plenaria, la moderadora abre la discusión para cada uno de los ítems llamando a un docente por vez para que comparta su respuesta a ¿Qué hizo que este ítem sea más difícil que el anterior?</li><li>- Una vez que se termina con la discusión de todos los ítems, la moderadora señalando las conclusiones y las ideas fuerza y recordando el sentido de la actividad.</li></ul>

---

<sup>20</sup> “El método bookmark, en términos generales, consiste en colocar marcas (tantas como cortes se hayan preestablecido) en un cuadernillo de ítems ordenado por dificultad (estando esta dificultad estimada mediante procedimientos de teoría de respuesta al ítem). La pregunta típica que guía el establecimiento de cortes es la siguiente: “¿Hasta qué ítem debe ser capaz de resolver un estudiante, como mínimo, para ser considerado parte del nivel?”. El procedimiento establece que los jueces, organizados en grupos pequeños, determinan, en tres rondas, los cortes para cada nivel de desempeño (UMC, 2020, p. 91)

(iii) Modelamiento de uso de ítemes liberados de las pruebas en actividades del aula

Una práctica que tiende a traer consigo las pruebas estandarizadas (generalmente en formato de opción múltiple) es la aplicación frecuente que hacen algunos docentes de pruebas iguales o similares con sus estudiantes. Estos docentes enfocan más su atención y tiempo en “preparar” a los estudiantes para la prueba que en implementar el currículo. Sin embargo, es difícil frenar esta práctica. De allí que una estrategia útil puede ser mostrarles a los docentes cómo pueden usar los ítemes liberados de las pruebas estandarizadas de una manera más reflexiva en la lógica de desarrollo de competencias.

### Cuadro 3

#### Modelamiento de uso de ítemes liberados de las pruebas en actividades del aula

Objetivo de la actividad	Aprovechar ítemes liberados de las pruebas para modelar sesiones en las cuales los docentes propicien el desarrollo de capacidades lectoras en sus estudiantes en el marco de una interacción de alta demanda cognitiva.
¿En qué consiste?	<ul style="list-style-type: none"><li>• En el grupo la moderadora explica a los participantes el objetivo de la actividad y cuál será la dinámica que se seguirá.</li><li>• Se distribuye a los docentes el ítem que se utilizará para que lo lean.</li><li>• Luego se presentan dos sesiones de clase en las que se utiliza el ítem (ver Anexo 4). Se puede representar a través de una dramatización o se puede usar un vídeo.</li><li>• Luego se pide analizar a los docentes cada uno de los casos y al final la moderadora puede cerrar enfatizando las siguientes ideas:<ul style="list-style-type: none"><li>• En el primer caso, el docente 1 solo verifica si es que los estudiantes han marcado la respuesta correcta. Solo hace eso. Con las otras preguntas hace lo mismo. Ni siquiera evalúa, pues no revisa o explica la fuente de los errores ni por qué tal o cual alternativa es la correcta.</li><li>• En el segundo caso, el docente 2 no se centra en la prueba. Utiliza el texto y las preguntas de esa prueba porque le parece interesante. En concreto lo que hace el docente es:<ul style="list-style-type: none"><li>○ Conversa sobre el texto libremente con sus alumnos. Rescata lo que les gusta o les interesa.</li><li>○ Plantea la pregunta, pero sin las alternativas. De esta manera, promueve la participación colectiva, desarrolla la expresión oral, la elaboración de ideas y el intercambio grupal.</li><li>○ Trabaja la fuente del error, reorienta el camino a la respuesta, analiza el texto.</li><li>○ Nuevamente trabaja el porqué del error en la pregunta con las alternativas. Un ítem está compuesto por una alternativa correcta y los demás son distractores (alternativas incorrectas). Cada vez que un niño o niña escoge un distractor, el profesor tiene la oportunidad de enseñarle por qué se pudo haber equivocado y reorientarlo. Se vuelve a leer partes del texto.</li><li>○ No se queda en la lectura del texto ni en la pregunta. Además de desarrollar la comprensión lectora y la expresión oral a través de una pregunta de la prueba, también se vislumbra el desarrollo de una mínima capacidad de investigación, de búsqueda de información distinta a la del texto propuesto, de extensión del tema</li></ul></li></ul></li></ul>

visto. Y ya queda esbozada, además, la próxima sesión relacionada a la producción de textos.

- Es más, no se queda solo en la prueba. Porque en las otras clases, trabaja otras actividades utilizando los cuadernos de trabajo, los libros de texto, varios textos que trae de fuera de la escuela y que encuentra en revistas, periódicos, enciclopedias, etc.
  - No necesita aplicar de manera frecuente evaluaciones tipo. Solo sabe que debe trabajar distintos tipos de texto y sobre temas variados.
- 

Fuente: Extraído de Llanos, 2016.

## **Uso de resultados para el diseño e implementación de políticas educativas**

La información brindada por las evaluaciones censales en Perú ha tenido un uso intensivo en los últimos años para programas de incentivos como los Compromisos de Desempeño,<sup>21</sup> que otorga recursos adicionales a las instancias de gestión educativa regional y local por el cumplimiento de determinadas metas educativas (por ejemplo, distribución oportuna de materiales educativos a las escuelas, mejora en los resultados en la ECE en su jurisdicción), o el Bono Escuela, que brinda un reconocimiento monetario a los docentes y directores de las escuelas estatales por el logro de resultados de sus estudiantes en la ECE.

Adicionalmente a ello los resultados se han utilizado como un insumo para la focalización y evaluación de impacto de las principales intervenciones del Ministerio, como el programa de Acompañamiento Pedagógico o el modelo secundario de Jornada Escolar Completa. En la misma dirección, la información de los resultados ha sido usada para el diseño y evaluación de algunas innovaciones costo efectivas que desarrolla el laboratorio de innovación MineduLAB<sup>22</sup>.

Como se puede observar los resultados de las evaluaciones se han centrado en el diseño o evaluación de mecanismos de incentivos y programas más que en políticas propiamente dichas. Se puede afirmar que, en el caso de Perú, los estándares de aprendizajes y en menor medida el currículo nacional han sido instrumentos de política que se han elaborado en base a las evidencias de las evaluaciones. Probablemente este es el uso más claro que han tenido la información brindada por las evaluaciones con relación al desarrollo de políticas educativas. Como se ha señalado en la sección 3 los resultados de las evaluaciones nacionales como internacionales sirvieron como un insumo esencial para la elaboración de los Mapas de Progreso del Aprendizaje, así como para el diseño del nuevo currículo aprobado en 2016. En el caso de los Mapas de Progreso la información de las evaluaciones se utilizó, entre otras cosas, para mejorar la cobertura de contenido del enunciado de los niveles del Mapa; mejorar la identidad del nivel u homogeneidad de su dificultad, de manera que se procure que, en un mismo nivel, no se incluyan logros que, según lo mostrado en las pruebas, pertenecen a niveles de dificultad diferentes; ajustar la exigencia que se considera apropiada para definir el estándar (Tapia, et al, 2016).

---

<sup>21</sup> <http://www.minedu.gob.pe/cdd/>

<sup>22</sup> <http://www.minedu.gob.pe/minedulab/cuarta-ventana.php>

## Referencias

- Bracho, T. (2022). Avances y retos del INEE en su etapa de organismo constitucional autónomo 2013-2019. En F. Miranda & A. Martínez (Eds.) *Políticas, reformas y evaluación educativas en México y América Latina: balance y perspectivas a futuro*. Universidad Iberoamericana.
- Burga León, A. (2019). Aplicaciones de la tecnología a la evaluación psicométrica. *Propósitos y Representaciones*, 7(SPE), e318.
- Dargent, E. (2014). Determinantes internacionales de la capacidad de las agencias estatales: lecciones a partir de Colombia y el Perú. *Apuntes: Revista de Ciencias Sociales*, 41(74), 9-40.
- Escuela Iberoamericana de Gobierno Educativo [EIGE] (2011) *¿Cómo devolver los resultados de las evaluaciones a las escuelas para lograr la mejora educativa?* Informe del Seminario realizado el 10 y 11 de noviembre de 2011 en Brasilia.
- Ferrer, G. & Fiszbein, A. (2015). *¿Qué ha sucedido con los sistemas de evaluación de aprendizajes en América Latina?* Lecciones de la última década de experiencia. El Diálogo. World Bank Group.
- Francis, S. (2023). *Análisis de la gobernanza y el proceso de macro-evaluación en Costa Rica: evolución, alcances y desafíos para promover la mejora del sistema educativo*. Investigación de base para el Informe Estado de la Educación 2022.
- Henríquez, C. (9-11 de agosto de 2016). *Uso de resultados de las evaluaciones y mejora escolar*. [Diapositivas de PowerPoint] Seminario Aprendizaje y Docencia en la Agenda de Educación 2030. Santiago.
- Henríquez, C. (2022). Avances y desafíos del sistema educativo y del Sistema Nacional de Evaluación en Chile. En F. Miranda & A. Martínez (Eds.) *Políticas, reformas y evaluación educativas en México y América Latina: balance y perspectivas a futuro*. Universidad Iberoamericana.
- Instituto Nacional para la Evaluación de la Educación [INEE] (2014). *Criterios para el desarrollo y uso de instrumentos de evaluación educativa 2014-2015*. INEE.
- Llanos, F. (2016) ¿Cómo usar las pruebas de lectura ECE en el aula o por qué podría desaparecer el mono pigmeo? *Educacion* 21.
- Ministerio de Educación de Perú [Minedu] (2017). *Currículo Nacional de la Educación Básica*. Minedu.
- Peri, A. (9-11 de agosto de 2016). *Una evaluación para la mejora de la Enseñanza: La evaluación formativa en línea*. [Diapositivas de PowerPoint] Seminario Aprendizaje y Docencia en la Agenda de Educación 2030. Santiago.
- OCDE (2013) *Synergies for better learning: an international perspective on evaluation and assessment*. *OECD Reviews of Evaluation and Assessment in Education*. OECD Publishing.

- Oficina de Medición de la Calidad de los Aprendizajes [UMC] (2020). *Reporte técnico de las Evaluaciones Censales y Muestrales de Estudiantes 2019*. Ministerio de Educación.
- Pinkasz, D. (2021) *El uso de los resultados de las pruebas de evaluación de los aprendizajes en el planeamiento de las políticas educativas en seis países de la región*. IIPÉ UNESCO.
- Ravela, P, Wolfe, R., Valverde, G., & Esquivel, J.M. (2006) Los próximos pasos: ¿Cómo avanzar en la evaluación de aprendizajes en América Latina? En: P. Arregui, *Sobre estándares y evaluaciones en América Latina*. PREAL.
- Ravela, P, Arregui, P., Valverde, G., Wolfe, R., Ferrer, G., Martínez Rizo, F., Aylwin, M. & Wolff, L. (2008) Las evaluaciones educativas que América Latina necesita. *Revista Iberoamericana de Evaluación Educativa*, 1(1), 51-63.
- Sempé, L. Andrade, P. (2017) *Evaluación de uso de informes de la Evaluación Censal de Estudiantes en la escuela*. Informe final. Proyecto FORGE.
- Tapia, J, Gysling, J. & Fernández, A. (2016) *Estándares de Aprendizajes como Mapas de Progreso: elaboración y desafíos*. El caso de Perú. SINEACE.
- UNESCO (2019) *The promise of large-scale learning assessment acknowledging limits to unlock opportunities*.
- Universidad del Pacífico & Centro de Medición (MIDE UC) (2017). *Diseño de una prueba estandarizada para la admisión a la educación superior*. Informe final.
- Yang, C., Luo, L., Vadillo, MA, Yu, R. y Shanks, DR. (2021). Las pruebas (cuestionarios) impulsan el aprendizaje en el aula: una revisión sistemática y metaanalítica. *Boletín Psicológico*, 147 (4), 399–435.

## **Anexos**

### **Anexo 1**

#### **Orientaciones de política para los sistemas de evaluación**

<b>Dominio</b>	<b>Orientaciones de política</b>
Gobernanza	<ul style="list-style-type: none"><li>• Establecer un marco coherente para la evaluación</li><li>• Desarrollar objetivos claros y progresiones de aprendizaje para orientar la evaluación</li><li>• Asegurar un buen balance entre evaluaciones sumativas y formativas</li><li>• Establecer salvaguardas para evitar una dependencia y sobre confianza de las evaluaciones estandarizadas</li><li>• Compartir responsabilidades para la gobernanza e implementación del sistema de evaluación</li></ul>
Procedimientos	<ul style="list-style-type: none"><li>• Delinear una variedad de tipos de evaluaciones para obtener una imagen más completa de los aprendizajes de los estudiantes</li><li>• Apoyar los procesos efectivos de evaluación formativa</li><li>• Clarificar e ilustrar criterios para juzgar el desempeño en relación con los objetivos nacionales</li><li>• Garantizar la coherencia y corrección de las evaluaciones</li><li>• Promover formatos de evaluación que capturen las competencias claves</li><li>• Basarse en enfoques innovadores desarrollados en sectores educativos particulares</li><li>• Aprovechar el potencial de las TIC para desarrollar instrumento de evaluación más complejos</li><li>• Garantizar que la evaluación sea inclusiva y sensible para las necesidades de todos los estudiantes</li></ul>
Capacidad	<ul style="list-style-type: none"><li>• Poner al estudiante en el centro y desarrollar la capacidad de los estudiantes para participar en su propia evaluación</li><li>• Mantener la centralidad de la evaluación basada en el docente y promover la capacidad profesional docente</li><li>• Identificar áreas prioritarias de evaluación para la formación inicial y el desarrollo profesional de los docentes</li><li>• Utilizar los procesos de valoración de docentes y de evaluación escolar para ayudar a los docentes a desarrollar sus capacidades de evaluación</li></ul>
Uso de resultados	<ul style="list-style-type: none"><li>• Desarrollar directrices claras para la presentación de informes</li><li>• Involucrar a los padres de familia en la educación a través de informes y comunicaciones adecuados.</li><li>• Garantizar la transparencia y la equidad al utilizar los resultados de la evaluación para decisiones de altas consecuencias</li><li>• Promover el uso regular de los resultados de la evaluación para la mejora</li></ul>

Fuente: OECD, 2013.

## **Anexo 2**

### **Atributos de la evaluación sumativa**

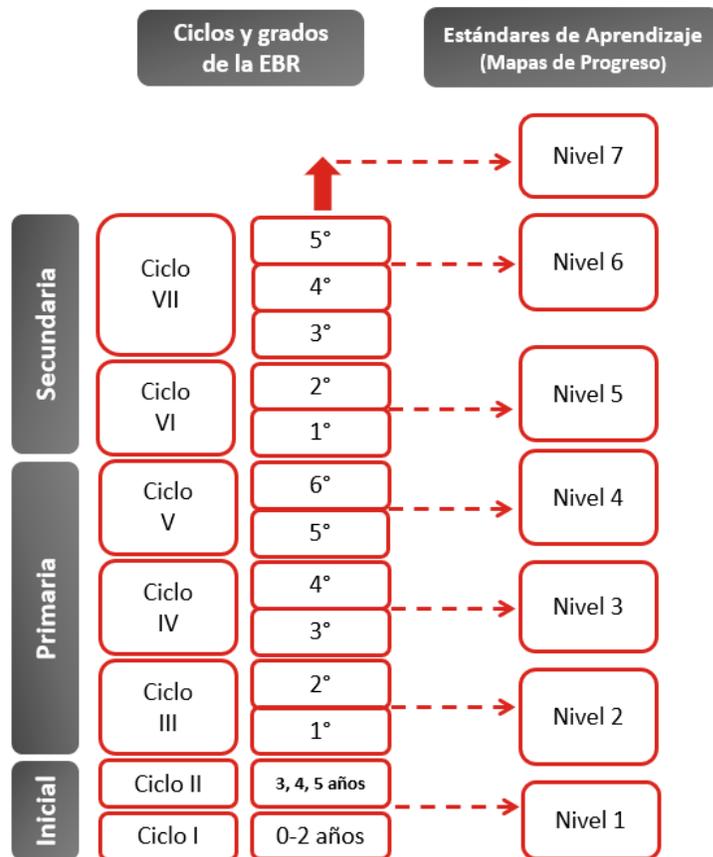
<b>Aspecto</b>	<b>Propuesta</b>	<b>Justificación</b>
Cobertura educación básica	<ul style="list-style-type: none"><li>• Evaluar grados intermedios y finales de ciclo de la educación básica.</li><li>• Se propone evaluar censalmente a grados finales de ciclo intermedios y evaluar muestralmente a grados/ciclos finales de cada nivel<sup>a/</sup></li></ul>	<ul style="list-style-type: none"><li>• Permite monitorear a los estudiantes en diferentes momentos de su trayectoria escolar</li><li>• La evaluación de grados intermedios moviliza a los centros educativos en tanto los estudiantes permanecen en él y están bajo su responsabilidad.</li><li>• Evita que los centros se preocupen sólo por la cohorte de estudiantes evaluados.</li></ul>
Cobertura curricular	<ul style="list-style-type: none"><li>• Evaluar competencias y contenidos curriculares más allá de lectura y matemática.</li><li>• Evaluar un módulo común de competencias instrumentales: lectura, matemática, ciudadanía, ciencias.</li><li>• Módulo variable: escritura, inglés, TIC, etc.</li></ul>	<ul style="list-style-type: none"><li>• Asegura la evaluación de competencias claves e instrumentales.</li><li>• Evita el estrechamiento curricular.</li></ul>
Cobertura poblacional	<ul style="list-style-type: none"><li>• Combinar evaluaciones censales y muestrales</li><li>• Incluir pruebas diversificadas para población con características específicas y que presentan desventajas sociales: discapacidad e indígena.</li></ul>	<ul style="list-style-type: none"><li>• Permite contar con el poder movilizador de las censales y amplía la información en áreas, grados en los que puede resultar inviable o innecesariamente censar.</li><li>• Reconoce la diversidad.</li></ul>
Periodicidad	<ul style="list-style-type: none"><li>• Dar estabilidad al sistema mediante la recurrencia en el esquema de evaluación.</li></ul>	<ul style="list-style-type: none"><li>• Permite establecer comparaciones en el tiempo (al evaluar periódicamente el mismo grado). Este atributo junto con las evaluaciones censales brinda información longitudinal para ciertas cohortes.</li><li>• Promueve un mayor uso de la información dado que los actores pueden anticipar sus cursos de acción.</li></ul>

a/ La excepción a esta propuesta sería la Prueba Nacional de bachillerato. Este tema se tratará de manera específica al final de esta sección.

Fuente: Elaboración propia.

### Anexo 3

#### Estructura de la Educación Básica Regular y Estándares de aprendizajes



Fuente: Minedu, 2017.

## **Anexo 4**

### **Detalle del Modelamiento de uso de ítemes liberados de las pruebas en actividades del aula**

#### **Sesión 1.**

Profesor 1: A ver, alumnos, leamos este texto de la Evaluación. Se llama “Un monito muy especial”. Les reparte a sus alumnos la fotocopia del texto y unas preguntas para marcar.

*Les hace leer en voz alta. Les corrige la pronunciación; que lean alto, que vocalicen, que entonen, que “respeten” la puntuación.*

Luego el profesor les dice:

- Bien, ahora respondamos las preguntas. La primera pregunta. Dice “¿Por qué podría desaparecer el mono pigmeo?”
- Los chicos miran la prueba
- ¿Es la a, es la b o es la c? – pregunta el profesor.
- Es la “a” –responden varios.
- No, es la “b” -responden otros.
- Profesor 1 ¡Chicos, es la b! Corrijan los que han marcado mal.
- Listo, siguiente pregunta.

#### **Sesión 2.**

Profesora 2: A ver, alumnos, leamos este texto. Se llama “Un monito muy especial”. Les reparte el texto. Solo el texto.

*Una vez que han terminado de leer el texto en silencio, plantea unas preguntas.*

Profesora 2: ¿Qué es lo que más les ha gustado o lo que más les ha llamado la atención de este texto?

*La profesora se pone a conversar sobre el texto. Aclaran ideas, comparten lo que más les ha interesado, explican algunas de ellas, atiende las preguntas de los niños y niñas, se ríen, rescatan emociones, las vinculan con algunos aspectos de su vida o con algo que han leído.*

Luego los niños y niñas comienzan a responder las preguntas de la prueba.

Pero así

- *La profesora escribe la pregunta en el papelote o pizarra. Solo la pregunta, sin alternativas (“a”, “b” o “c”).*

*En la pizarra queda escrito: “¿Por qué podría desaparecer el mono pigmeo?”.*

*Algunos estudiantes comienzan a participar:*

Alberto: “Va a desaparecer porque es muy débil”

Julieta: “Los monos pigmeos desaparecen porque los matan”

*Son niños que usaron algún tipo de experiencia o saber previo, pero no los ha ajustado, contrastado, conectado con el texto que acaban de leer. Sus respuestas son incorrectas.*

- Profesor 2: Chicos, chicas, es cierto, algunos monos pigmeos o titís se pueden morir porque los matan o porque están débiles. Pero según el autor del texto, los titís pueden desaparecer por otra razón. ¿Cuál es la razón que menciona el autor? ¿encontramos esa parte del texto sobre la desaparición de los monos pigmeos?

Pedro: Es el último párrafo

Lucía: Sí, porque destruyen los árboles. Por eso desaparecen los monos pigmeos

Profesor 2: Pero qué tiene que ver los árboles con su extinción de los monos titís

Lucía: Porque ahí viven y si no tienen dónde vivir, se van a morir.

Profesor 2: Muy bien, chicos, chicas.

*Más adelante, les reparte la misma pregunta y sus alternativas:*

¿Por qué podría desaparecer el tití pigmeo?

- a. Porque es un animal demasiado pequeño
- b. Porque están destruyendo el lugar donde viven
- c. Porque mucha gente se lo lleva como mascota

¿Cuál será la alternativa correcta?

La “b” dicen muchos, pero Hilda dice que es la a).

El profesor 2 le pregunta a Hilda por qué cree que el mono pigmeo podría desaparecer

Hilda: Porque lo dice el texto, arriba, cuando comienza. Mira. Ahí está. Dice que es el mono más pequeño del mundo. Y si algo es chiquito, pues, lo aplastan, se lo comen otros animales.

Profesor 2: Sí, dice que es el más pequeño del mundo, pero no hay nada en el texto que haga suponer que por eso desaparecen ¿no? Volvamos a leer ese primer párrafo.

*Hilda lo lee. Asiente*

- A ver Hilda, tú misma, dale una leída al último párrafo.

Hilda revisa el último párrafo.

Hilda: Ah ya, claro, porque cortan los árboles. Ahí viven los monos titís y después no tienen dónde vivir y desaparecen, claro.

Profesor 2: Entonces ¿cuál marcarías?

Hilda: La b)

Profesor 2: Entonces los monos pigmeos desaparecerían porque el hombre destruye los lugares donde viven.

Al final de clase, el profesor 2 dice:

- Chicos, para la siguiente clase, por favor, averigüen sobre un animal que esté en peligro de desaparición y averigüen por qué podría extinguirse. Traigan en un papel la información, solo datos, como en borrador, nada más.

Aquí en clase, formaremos grupos y trabajaremos la producción escrita de un texto con la información que traigan.