

CAPÍTULO

EDUCACIÓN PREESCOLAR, GENERAL BÁSICA Y DIVERSIFICADA

3

Desafíos de la Macroevaluación en el Sistema Educativo Costarricense

HALLAZGOS RELEVANTES

- El sistema de macroevaluación educativa de Costa Rica se ha reducido a un sistema de medición del logro de aprendizajes y presenta desafíos estratégicos para interrelacionar otros elementos clave de su funcionamiento, como el marco teórico-normativo, la organización institucional y los mecanismos de difusión y uso de resultados.
- El escaso uso de los resultados constituye un vacío histórico del sistema. Existe una falta de capacidades administrativas y técnicas que permitan transformar los resultados en información y conocimiento para el mejorar la calidad y equidad educativa.
- El marco teórico y normativo de la macroevaluación cuenta con alcance limitado: solamente se contempla a los estudiantes como objetos de evaluación, además de que no precisa cuál debería ser la participación de los diferentes actores de la comunidad educativa en el análisis de los resultados y en la implementación de soluciones.
- La organización institucional a cargo de la macroevaluación ha estado liderada por el MEP. Dicha organización ha presentado cambios recurrentes en las dependencias o entidades a cargo. Las funciones se han asociado más a los procesos de logística de la implementación de pruebas y elaboración de informes que a la promoción del uso de los resultados para la toma de decisiones.
- La institucionalidad a cargo de la macroevaluación educativa opera con conflictos de interés en la política pública, pues la instancia encargada de la ejecución de esta política (MEP) es también la encargada de evaluar sus resultados.
- A partir del 2021, la inversión en gestión y evaluación de la calidad se ha reducido. El 2023 presentó un decrecimiento del 23%: el más significativo de la última década.
- La macromedición de los aprendizajes ha experimentado, durante la última década, un retroceso significativo caracterizado por periodos de inestabilidad y cuestionamientos técnicos que no se han consolidado en instrumentos robustos.
- Desde 2012, las pruebas estandarizadas de Matemáticas no han conseguido articularse con los programas de estudio vigentes, lo cual ha obstaculizado la generación de insumos fundamentales para fortalecer la práctica docente y la toma de decisiones prontas y pertinentes para mejorar la acción en el aula.
- El país está lejos de implementar las prácticas más efectivas llevadas a cabo en el ámbito internacional, como las siguientes: autonomía técnica en la organización institucional del sistema, marco teórico y normativo con equilibrios entre la evaluación sumativa y formativa, articulación entre pruebas estandarizadas y los programas de estudio y empleo de estrategias difundir masivamente sus resultados y mejorar las acciones que se realizan en aulas y en los centros educativos.

CAPÍTULO

EDUCACIÓN PREESCOLAR, GENERAL BÁSICA Y DIVERSIFICADA

3

Desafíos de la Macroevaluación en el Sistema Educativo Costarricense

ÍNDICE	Introducción	Argumento central
Hallazgos relevantes 161	<p>Este capítulo analiza la situación actual y los desafíos que enfrenta Costa Rica para contar con un sistema de macroevaluación educativa apegado a las prácticas internacionales más efectivas en la materia.</p> <p>El análisis utiliza como referencia conceptual el término de macroevaluación planteado por la OCDE (2013), definido como el conjunto de factores, procesos, instrumentos y metodologías que operan con un enfoque sistémico en el que cada elemento está estrecha y coherentemente articulado con los objetivos de aprendizaje propuestos en la política curricular.</p> <p>El aspecto central de esta definición, que es de naturaleza tanto analítica como normativa, reside en afirmar que la macroevaluación constituye un sistema. Esto permite examinar la evolución y estado de sus diversos componentes y establecer parámetros de comparación con las experiencias internacionales. Aunque esta perspectiva incluye también al personal docente, el texto se centra en la evaluación aplicada a personas estudiantes. Las razones para esta escogencia son dos: por una parte, con esta población se ha desarrollado más experiencia evaluativa en Costa Rica; por otra, existe una necesidad de desarrollar nuevos enfoques que ayuden a revertir la pobreza de aprendizajes, la cual se ha visto severamente agudizada por el apagón educativo.</p>	<p>La macroevaluación, entendida como un sistema, representa un componente medular del proceso educativo. Su alcance es mucho más amplio que la aplicación de pruebas nacionales a estudiantes. Este argumento se desarrolla en cinco ideas fundamentales:</p> <ul style="list-style-type: none"> • Un sistema de macroevaluación comprende cuatro componentes indispensables relacionados entre sí: el marco teórico y normativo, la institucionalidad responsable de la evaluación, los instrumentos de medición y la difusión y usos que se da a los resultados. • El país no ha logrado consolidar un sistema robusto, transparente y efectivo, sino que se ha limitado a desarrollar la medición de aprendizajes con poco uso de los resultados para promover la mejora educativa. • La institucionalidad a cargo de la macroevaluación ha enfrentado constantes tensiones políticas y conflictos de interés que han propiciado una limitada innovación, cambios abruptos sin aprendizaje y rigideces para articular la evaluación con la política curricular. • Los instrumentos de medición en el país han experimentado un retroceso significativo en la última década en la que se han estancado o deteriorado los aprendizajes estudiantiles. • Costa Rica está lejos de adoptar las prácticas utilizadas por los referentes internacionales en la materia, lo que limita su eficiencia en la utilización de los recursos y en el uso de información para el diseño de políticas basadas en evidencia. • En el contexto de la crisis educativa actual, contar con un sistema de macroevaluación de los aprendizajes técnicamente robusto es necesario para generar información clave que permita diseñar estrategias de recuperación.
Introducción 163		
Estudios previos sobre macroevaluación en Costa Rica carecen de un análisis sistémico 164		
Más allá de la medición: importancia y visión sistémica de la macroevaluación educativa 165		
Desafíos y situación actual de la macroevaluación en Costa Rica 169		
Costa Rica lejos de implementar prácticas efectivas de macroevaluación educativa 186		
Retos clave y portafolio de sugerencias para consolidar un sistema efectivo de macroevaluación educativa en Costa Rica 193		

La literatura sobre la macroevaluación desde una perspectiva sistémica como la que aborda este capítulo es muy escasa en Costa Rica pese a la gran urgencia y a las necesidades del país en esta área. En este sentido, el texto procura ser un insumo relevante y oportuno para la deliberación pública, así como para orientar la toma de decisiones en un contexto caracterizado por la profunda crisis educativa que las últimas ediciones del Informe han examinado.

Para la elaboración del capítulo, se recurrió a una red de expertos nacionales e internacionales en macroevaluación educativa, que generaron nueve investigaciones especializadas. Estos estudios fueron posibles gracias al Fondo Concursable del Estado de la Educación del Conare y el apoyo de la Fundación Yamuni Tabush. Todos estos documentos se encuentran a disposición del público en el sitio web www.estadonacion.or.cr.

El capítulo se organiza en cuatro apartados principales. El primero examina la importancia y configuración de los componentes del sistema de macroevaluación educativa arriba descritos. El segundo apartado aborda su evolución histórica, la situación actual y sus desafíos, con un énfasis especial (pero no exclusivo) en la asignatura de Matemática. La tercera sección ofrece una revisión de las prácticas internacionales y nacionales que implementan los países que han consolidado sistemas eficientes y se destacan las principales lecciones aprendidas que Costa Rica puede aprovechar. Finalmente, el último apartado tiene un carácter propositivo y ofrece un menú de recomendaciones para que los actores educativos y políticos cuenten con opciones para deliberar acerca de la mejor manera de resolver una serie de problemas concretos relativos a la macroevaluación en Costa Rica. Estas recomendaciones no pretenden ser una “receta”, sino más bien un portafolio de ideas por considerar surgidas de la experiencia nacional e internacional.

Estudios previos sobre macroevaluación en Costa Rica carecen de un análisis sistémico

Los estudios sobre la macroevaluación en Costa Rica forman un cuerpo de

literatura aún incipiente centrado sobre todo en el análisis e impacto de las distintas pruebas nacionales que se han aplicado en el país, pero no así en un análisis sistemático que aborde sus componentes fundamentales. Dicha omisión constituye un sensible vacío para valorar si esta organización evaluativa nacional es robusta o si, por el contrario, presenta importantes debilidades que ayudan a explicar varios de los problemas identificados por la literatura existente.

Algunas investigaciones han analizado los modelos de macroevaluación que el país ha adoptado a lo largo de su historia, principalmente bajo el esquema de las Pruebas Nacionales de Bachillerato (PNB), antecesoras. Dichas evaluaciones se complementan con otras pruebas estandarizadas internacionales en las que el país participa desde hace ya varias décadas, como por ejemplo las pruebas del Programa Internacional de Evaluación de Alumnos (PISA, por sus siglas en inglés) de la OCDE para estudiantes de 15 años y los estudios del Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (Llece) desarrollados por la Unesco en América Latina para primaria tales como Terce 2013 y Erce 2019.

En sucesivas ediciones del *Informe Estado de la Educación* se han analizado los factores que inciden en el rendimiento de los estudiantes a través de pruebas estandarizadas. Entre otros, se ha identificado que el aula, el contexto de la persona estudiante, las actitudes del estudiantado y sus relaciones interpersonales influyen en el desempeño. En las ediciones sexta y séptima del Informe, se realizaron estudios de observación en el aula que evidenciaron una falta de aprovechamiento de los resultados en las pruebas para mejorar el trabajo docente, así como una brecha entre la propuesta curricular y las actividades en el aula. Además, las Pruebas Nacionales de Bachillerato (PNB) no se ajustaban a los objetivos planteados en los programas de estudios reformados bajo un enfoque de habilidades. La octava edición del Informe analizó los determinantes causales del desempeño en áreas específicas (PEN, 2015, 2017 y 2021).

Otros estudios nacionales también

han señalado problemas relevantes por atender en cuanto a la organización y los instrumentos de la macroevaluación. Guzmán et al. (2006) evaluaron los procesos de macroevaluación en el sistema educativo de Costa Rica entre 1994 y 2004. Se encontró que había falta de conocimiento de los resultados por parte de los actores educativos, nulos mecanismos de retroalimentación y carencia de un sistema de información y divulgación nacional. También se evidenció poca coherencia entre el uso de los resultados y los propósitos de la política educativa, así como una falta de claridad en los objetivos y funciones de la micro y macroevaluación¹ y una falta tanto de acompañamiento como de ajuste entre los resultados y el contexto de la población estudiantil. Los autores comentan que los resultados se emplean principalmente para la promoción de estudiantes y rara vez se usan para implementar planes de mejoramiento.

Por su parte, Esquivel et al. (2006) realizaron una evaluación de las Pruebas Nacionales de Bachillerato de la educación formal en la que se analizó la congruencia entre los objetivos de dichas pruebas y los fines de la educación costarricense. Encontraron un mayor nivel de congruencia con el objetivo de contribuir a la formación integral de los estudiantes, pero menor en los objetivos de motivar a los padres y madres de familia para integrarse al apoyar el proceso educativo y el de lograr que los resultados obtenidos se constituyeran en una herramienta para supervisar el progreso del rendimiento escolar.

Asimismo, hallaron que el sustento teórico de las pruebas carecía de un posicionamiento claro sobre los conceptos básicos de medición y evaluación educativa, la existencia de escalas de calificación débiles para los procesos de construir, validar y calificar las pruebas, poca operacionalización del currículo y una reducida articulación de los temarios de referencia con respecto a los programas de estudio. En cuanto a la interpretación y uso de los resultados de las pruebas, el estudio encontró que las interpretaciones que se realizan son limitadas y parten de un enfoque más sumativo relacionado

con la promoción de la población estudiantil. Además, los informes enviados por las autoridades se realizan con tardanza, sin toma de decisiones por parte de las entidades correspondientes del MEP y con un inexistente enlace entre las direcciones de Desarrollo Curricular y Control de Calidad.

En otra investigación elaborada por Ruiz (2018), se reconoce la importancia de contar con un sistema de evaluación de alta calidad y se propone para ello un modelo detallado que se vincule a los objetivos, habilidades y competencias que señala el Programa de Matemáticas vigente desde 2012. De acuerdo con el autor, la implementación de un modelo evaluativo congruente con el currículo es una herramienta invaluable e indispensable para que los docentes mejoren la calidad de la enseñanza.

Por último, el estudio de González et al. (2018) analizó el modelo evaluativo de las pruebas estandarizadas. Los autores destacaron la urgencia de revisar en profundidad los objetivos de las PNB, las cuales tenían más de treinta años de antigüedad y solo habían sido modificadas una vez en 2018. En consonancia con estudios previos, identificaron la ausencia generalizada de divulgación y retroalimentación de resultados a nivel nacional, sin seguimiento alguno ni planes de mejora.

Con base en estos antecedentes, en el presente capítulo se retoma el análisis de la macroevaluación en el sistema educativo costarricense desde una perspectiva sistémica. Se considera, por ejemplo, que mejorar y ajustar las estrategias de enseñanza en las aulas a partir del análisis eficaz de los resultados de la macroevaluación por parte de los distintos actores educativos es un mecanismo clave para enfrentar la crisis educativa actual.

Más allá de la medición: importancia y visión sistémica de la macroevaluación educativa

La evaluación del sistema educativo no se limita al logro escolar, sino que además abarca la interrelación entre diversos actores y factores como los objetivos, el cuerpo docente y directivo, el contex-

to familiar y personal del estudiante, entre otros. Se trata, entonces, de un proceso complejo y académico con funciones transcendentales, entre las que se encuentran el monitorear y evaluar el sistema nacional de educación, establecer un sistema de rendición de cuentas, proveer información pública y actuar como proceso diagnóstico y formativo.

Este apartado tiene como propósito brindar las nociones conceptuales indispensables para comprender la relevancia del sistema macroevaluativo en los procesos de transformación educativa, así como la complejidad que implica su funcionamiento, al tratarse de un proceso sistemático, multifactorial y altamente especializado. En este sentido, se parte de la premisa de que los sistemas de macroevaluación educativa no se pueden limitar exclusivamente a la medición del logro educativo de los estudiantes. Aunque las pruebas estandarizadas y la medición de aprendizajes desempeñan un papel estratégico, reducir el sistema a su aplicación corresponde a una visión simplista y reducida que impide lograr el objetivo final: mejorar los aprendizajes de los estudiantes. Por lo tanto, a continuación, se aportan elementos para comprender la importancia de los procesos de macroevaluación educativa, cómo se conforman y organizan estos dentro de un sistema, el papel que desempeñan las pruebas estandarizadas y las interacciones necesarias para alcanzar su objetivo final.

La macroevaluación como factor determinante para la mejora de los aprendizajes

La macroevaluación es esencial para la transformación y la mejora de los sistemas educativos, ya que proporciona información objetiva y relevante para mejorar los procesos, medir la calidad y el progreso a lo largo del tiempo y establecer planes de mejoramiento. Además, según los aportes más recientes de la neurociencia, contribuye con el desarrollo de habilidades que preparan al estudiante para una vida exitosa, pues mejora la motivación, la memoria a largo plazo, la autorregulación y la confianza. Por lo tanto, la evaluación ocupa un lugar

central en el proceso educativo al proporcionar datos y análisis válidos y fiables para diagnosticar la situación del sistema y para tomar decisiones que permitan generar procesos de cambio y mejorar la calidad educativa (Guzmán et al., 2006; Tedesco, 2016; Cruz y Quiñones, 2012; Tiana, 2008).

En la actualidad, la evaluación educativa cobra más relevancia que nunca debido a las presiones para el cambio que experimentan los sistemas educativos nacionales, así como las crecientes demandas ciudadanas de información sobre los procesos educativos, la transparencia y la rendición de cuentas (OCDE, 2013). Bien utilizada, contribuye a mejorar la gestión del sistema educativo, la calidad de la enseñanza y la definición de políticas educativas nacionales, gracias a dos tipos de aportes principales: en primer lugar, visibilizar los resultados y los aspectos fundamentales de la labor educativa y, en segundo lugar, aportar información sobre los actores del sistema con diversos propósitos como certificar aprendizajes y señalar estándares de desempeños esperados (figura 3.1).

La mayoría de los países han centrado los procesos de macroevaluación en reportar de manera agregada² los resultados obtenidos principalmente en los aprendizajes de los estudiantes. Es decir, se suele omitir el hecho de que se trata de un proceso multifactorial enfocado no exclusivamente en el logro escolar y que abarca diversos objetos de estudio, actores institucionales (ministerios y secretarías de educación, o entidades independientes), alcances, objetivos, técnicas e instrumentos de medición y las formas en que se utilizan sus resultados para mejorar su objeto de estudio — como el desempeño estudiantil, docente o del centro escolar dentro del sistema educativo (Unesco, 1997)—.

A pesar de sus bondades, las evaluaciones nacionales de conocimientos no están exentas de limitaciones:

- Aportan información valiosa sobre la calidad de la educación, pero no son indicadores completos de esta, ya sea porque algunos saberes y aprendizajes

Figura 3.1

Principales aportes de las macroevaluaciones educativas estandarizadas al sistema educativo



Fuente: Vargas, 2023 con base en Rodino, 2023a.

- resultan difíciles de medir de esta manera o porque son específicos de un contexto local.
- La macroevaluación no es el único tipo de evaluación relevante. Esta puede complementarse con otras evaluaciones de menor escala del desempeño estudiantil, docente, institucional, de las políticas educativas, del uso de los recursos o curriculares, entre otras.
- Para ser útiles, deben concebirse, percibirse y emplearse como un mecanismo de responsabilidad pública de todos los actores vinculados al quehacer educativo. Si no se adoptan medidas correctivas o la solución se traslada a actores privados, no tendrán un efecto sistémico relevante.

- Resultan necesarias, pero no suficientes para mejorar la educación. Por ello, deberían estar alineadas con otros aspectos del sistema educativo

De acuerdo con Rodino (2023a), las evaluaciones deben responder a objetivos y necesidades claras y contar con los instrumentos adecuados para implementarlas e interpretar sus resultados. Además, esta autora las clasifica en dos grandes grupos: según los objetivos para los que se desarrollan y según su alcance más amplio o específico (figura 3.2). Esta agrupación no es excluyente: las evaluaciones concebidas para atender distintos objetivos pueden tener diversos alcances y las que tienen un determinado alcance puede estar asociada a diversos objetivos. Tampoco es excluyente a lo interno de cada clasificación, pues pueden realizarse evaluaciones diagnósticas, otras

formativas y sumativas, así como combinar macroevaluaciones con microevaluaciones.

En cuanto al ámbito de aplicación, las evaluaciones se pueden clasificar como macroevaluación o microevaluación. La primera se aplica a nivel país mediante el empleo de instrumentos o pruebas estandarizadas. La segunda se desarrolla en contextos más reducidos, como un aula o una escuela, y a través de instrumentos específicos para esa situación, normalmente diseñados e implementados por cada docente

La evaluación sumativa busca sintetizar el aprendizaje logrado por una persona a fin de llevar un registro, calificarlo o certificar sus logros. Sus resultados pueden utilizarse para expedir un documento oficial por parte de una autoridad educativa que certifica o acredita el nivel alcanzado. En cambio, la evaluación formativa mide lo que una persona está aprendiendo en momentos determinados de su proceso. Se emplea para identificar aspectos en los que resulta necesario profundizar y así orientar los siguientes pasos. Las evaluaciones diagnósticas³ corresponden a un subtipo de evaluación formativa que permiten identificar la base o el punto de partida sobre el cual guiar el aprendizaje para implementar las adecuaciones necesarias en la mediación pedagógica. Tanto la evaluación formativa como la diagnóstica requieren de un diseño cuidadoso, razón por la cual se amplía con más detalle en el recuadro 3.1.

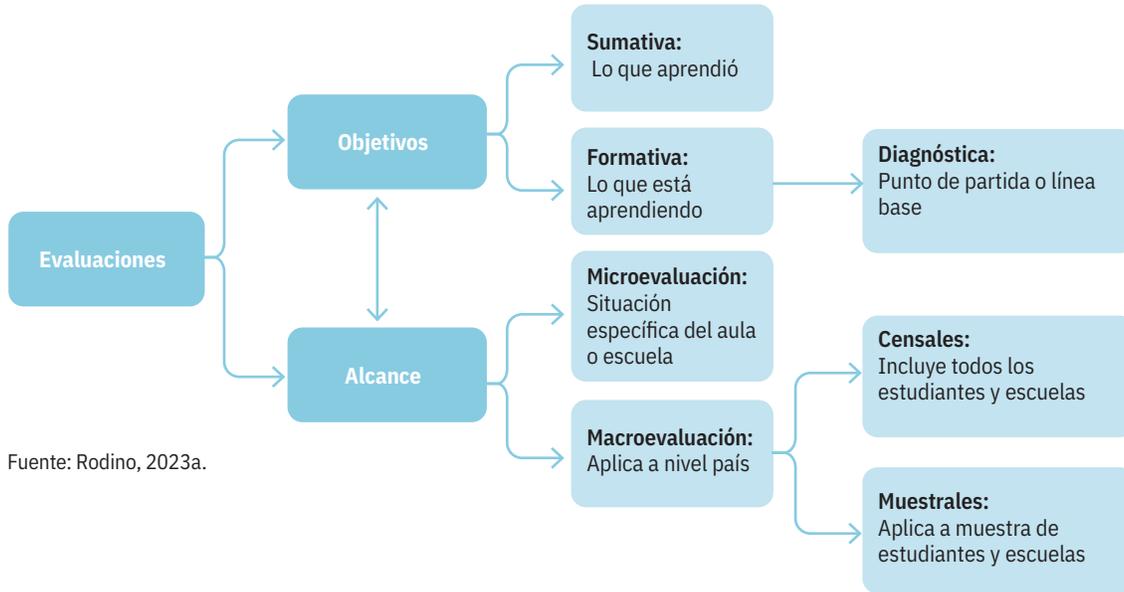
La necesidad de emplear un enfoque sistémico para evaluar los aprendizajes

A partir de una amplia revisión de literatura y el criterio de expertos internacionales (OCDE, 2013; Rodino, 2023a; Lasida, 2023; Medeiros, 2023; Ospina y Angulo, 2023; Miranda, 2023; Miranda y Monroy, 2023), se identificaron cuatro elementos clave que conforman un sistema de macroevaluación educativa (figura 3.3): el marco teórico y normativo, la institucionalidad, los instrumentos de medición y la difusión y usos de los resultados o insumos que provee el sistema.

El marco teórico y normativo, también conocido como marco de referencia, se

Figura 3.2

Tipos de evaluaciones educativas según su finalidad y ámbito de aplicación



Fuente: Rodino, 2023a.

Recuadro 3.1

La evaluación diagnóstica y formativa de aula requieren de un diseño cuidadoso con herramientas de apoyo a los docentes para que estos la apliquen con éxito y les sea útil.

La evaluación diagnóstica y formativa en el aula corresponde a un tipo de evaluación que tiene como objetivo principal que el docente obtenga información valiosa y puntual de los aprendizajes alcanzados por sus estudiantes. Además, procura identificar los riesgos de fracaso que estos presentan, así como ofrecer apoyo mediante estrategias específicas y personalizadas que les permitan una mejora educativa sustantiva. La evaluación diagnóstica se aplica al inicio del ciclo escolar y para que sea exitosa requiere una preparación esmerada y detallada de distintos instrumentos tanto en la aplicación, como en la devolución de los resultados y en el seguimiento de los estudiantes.

Una buena práctica en la aplicación de este tipo de pruebas ha sido desarrollada en los últimos años por el Mejoredu de México. En la fase de preparación y elaboración de instrumentos para cada una de las etapas señaladas participan especialistas, docentes, directores, supervisores y asesores técnico-pedagógicos, responsables del diseño curricular, así como investigadores y académicos en las distintas áreas de evaluación (Mejoredu, 2023). Destacan entre esos instrumentos las tablas descriptivas, tablas argumentativas, unidades de análisis, rúbricas para evaluar preguntas abiertas y orientaciones didácticas para el uso formativo de los resultados de las evaluaciones de los aprendizajes.

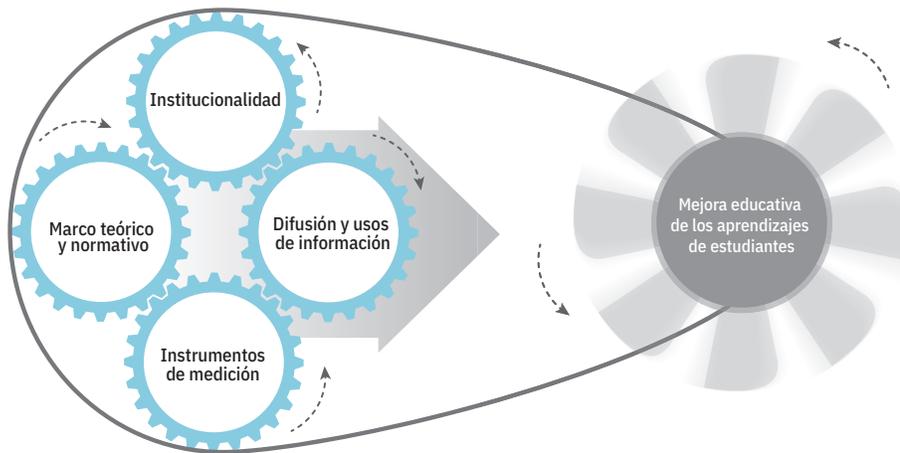
Una vez obtenidos los resultados es fundamental que cada centro educativo y sus comunidades generen una serie de procesos y espacios reflexivos para que el profesorado pueda analizarlos, compartirlos, discutirlos con sus colegas y comunicarlos tanto a los y las estudiantes como a sus familias.

Finalmente, para el seguimiento de los aprendizajes de los estudiantes es fundamental que el personal docente pueda apropiarse de los instrumentos y contar con otros adicionales que le permitan aplicar con éxito las estrategias de apoyo a cada estudiante y valorar cambios en un período determinado mediante la selección de las evidencias más sólidas y adecuadas.

Fuente: Mejoredu, 2023.

Figura 3.3

Sistema de macroevaluación educativa: componentes clave e interrelaciones



Fuente: Alfaro y Barquero, 2023.

refiere a las normativas que regulan y orientan la evaluación educativa a nivel macro, es decir, a nivel nacional o regional. Su conceptualización condiciona el funcionamiento y alcance del sistema. Dicho marco incluye leyes, reglamentos, políticas y directrices que establecen los objetivos, principios, criterios, estándares, modelos de evaluación y seguimiento y definición de componentes (evaluación a docentes, estudiantes y centros educativos). También comprende los procedimientos para evaluar el desempeño de los sistemas educativos, a saber: la conceptualización, los propósitos y los alcances del sistema, así como la coherencia con los objetivos de aprendizaje planteados en el currículum nacional y la articulación conceptual con los otros componentes. Para lograrlo, debe haber consenso entre los distintos actores del sistema educativo respecto a cuál debería ser el propósito del sistema de evaluación de aprendizajes.

Por su parte, a institucionalidad se refiere al conjunto de procesos y estructuras que rigen la planificación, la implementación, el seguimiento y la administración y uso de los recursos. Es decir,

establece la distribución de las potestades y capacidades para tomar decisiones e implementarlas. En consecuencia, la institucionalidad determina la operacionalización del sistema mientras que, a su vez, es responsable de asegurar la transparencia, la eficacia, la legitimidad del proceso y garantizar el uso eficiente de sus recursos. Las responsabilidades de su funcionamiento pueden recaer sobre uno o varios actores, pero, como se verá más adelante en este capítulo, la evidencia muestra que los países con estructuras de gobernanza con mayor autonomía administrativa y técnica han logrado innovar, ampliar sus capacidades y desarrollar experiencias técnicas sólidas y estables.

Los instrumentos de medición, por otra parte, son el componente operativo, los medios que permiten realizar las evaluaciones y, por esta vía, constituyen las herramientas de gestión educativa que ofrecen la posibilidad de valorar el cumplimiento de los programas de estudio definidos en la política curricular. La literatura los reconoce, en teoría, como un recurso educativo habilitador de capital social porque la información que de

ellos se desprende genera conocimiento, permite monitorear y comprender las brechas educativas, identificar tanto las fortalezas como las debilidades de cada actor educativo y orientar el uso eficiente de los recursos a partir del diseño de estrategias educativas diferenciadas.

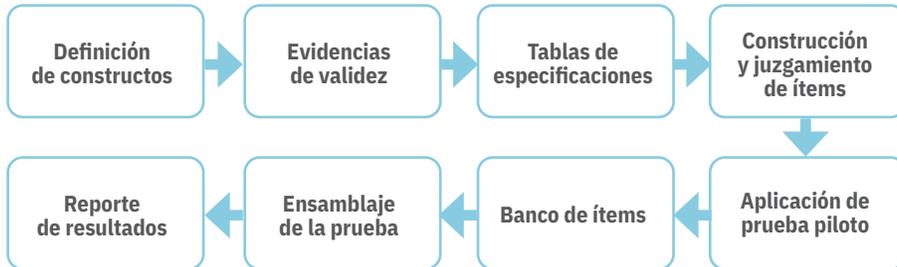
Cabe señalar que en las macroevaluaciones se emplean pruebas estandarizadas o uniformes para cada uno de los actores educativos que se pretende evaluar. Debido a su amplio alcance, estos instrumentos deberían ser diseñados y validados rigurosamente por especialistas en psicometría siguiendo estrictos protocolos para garantizar que las puntuaciones sean comparables y que las inferencias sobre las personas examinadas sean válidas (Rodino, 2023a). En este sentido, resulta esencial asegurar la legitimidad y la rigurosidad técnica de los instrumentos utilizados, especialmente en pruebas de alta consecuencia o de certificación para las personas examinadas, a fin de garantizar una alta credibilidad de las interpretaciones y de los resultados derivados de su aplicación. Según Moreira et al. (2022) la elaboración de pruebas estandarizadas se conforma de varias etapas que se detallan adelante (figura 3.4).

El cuarto componente clave de la macroevaluación educativa corresponde a la difusión y el uso de los resultados que se generan a partir de las mediciones aplicadas a los distintos componentes de la evaluación. Esto no solo depende de la información de calidad derivada de las mediciones estandarizadas, sino que están condicionados por las decisiones de la gobernanza ministerial y, por consiguiente, del diseño, ejecución y seguimiento de las intervenciones educativas que se lleven a cabo.

Desde un punto de vista normativo, la interacción entre los cuatro componentes debería facilitar el cumplimiento del objetivo final del sistema de macroevaluación, a saber: la promoción de la mejora educativa continua (figura 3.3). Sin embargo, esta mejora depende de la existencia de claridad en los objetivos del marco de referencia, así como de una estructura de gobernanza dotada de competencias y capacidades para implementar la política educativa, de las formas en que se

Figura 3.4

Características esenciales del procedimiento técnico para la elaboración de pruebas estandarizadas



Fuente: Barquero, 2023 con base en Moreira et al., 2022.

presenten los resultados, la información que se comparta y, crucialmente, del hecho de que los hallazgos sean interiorizados por las escuelas, docentes, administrativos padres de familia y la sociedad civil en general (Ferrer y Fiszben, 2015).

Bajo estas condiciones, es posible concebir la macroevaluación como un sólido sistema de gestión educativa que informa sobre el desarrollo de políticas educativas, colabora en la implementación y el alcance de los planes de estudio, propone estrategias de mejora, asignaciones presupuestarias y guía en las decisiones de asignación de recursos, en la gestión del desempeño académico y en la identificación de brechas educativas.

Finalmente, cabe señalar que un sistema de macroevaluación no es un fin en sí mismo; más bien, se constituye como un componente integrador de los procesos globales de la educación y comprende las distintas instancias y actores que participan en el proceso de enseñanza-aprendizaje en todos sus niveles. La literatura advierte que los procesos educativos no deben funcionar como componentes independientes, pues, de lo contrario, se obstruye su potencial (OCDE, 2013). La ausencia de alguno de sus componentes o la desconexión entre ellos deja abierto un espacio para generar desconfianza en el sistema y cuestionamientos a su relevancia a la hora de promover mejoras de los aprendizajes.

En suma, para lograr impactos y mejoras en los aprendizajes y en la calidad, es necesario armonizar las políticas con los

instrumentos de evaluación, acotar sus propósitos con las apuestas nacionales, mejorar los ambientes de aprendizaje y definir un portafolio de acciones y programas que permitan emprender acciones de mejora continua. De esta forma, se logrará empoderar a los distintos actores de la comunidad educativa, tanto en los establecimientos educativos, como en las aulas (Ospina y Ángulo, 2023).

Desafíos y situación actual de la macroevaluación en Costa Rica

Este apartado del capítulo analiza la situación actual de la macroevaluación educativa en Costa Rica desde una perspectiva histórica. Además, aborda los desafíos que enfrenta el sistema educativo ante el reto de contar con un sistema nacional de macroevaluación robusto, transparente y efectivo como herramienta fundamental para promover la mejora educativa de los aprendizajes. El análisis se realiza a partir de los componentes clave del sistema presentados en la sección previa, a saber: los marcos de referencia de la evaluación de aprendizajes, la gobernanza, las técnicas y métodos de medición de aprendizajes y la difusión y uso de resultados.

Como se verá, la principal conclusión de este apartado es que la estructura actual de la macroevaluación en el país presenta vacíos y limitaciones en todos los componentes, los cuales, además, funcionan de manera aislada o poco

articulada entre ellos, lo que impide su consolidación como un sistema robusto, independiente y confiable. Asimismo, los principales hallazgos señalan que, este sistema, desde una perspectiva histórica, se ha reducido a la medición de aprendizajes de los estudiantes, la cual presenta deficiencias al utilizar, difundir la información y generar análisis nacionales que permitan mejorar las prácticas de aulas y el desempeño educativo.

En cuanto sus componentes, se observa que el marco teórico y normativo se limita a la implementación de pruebas nacionales en la población estudiantil y presenta carencias al definir la participación de otros actores educativos. La gobernanza del sistema macroevaluativo costarricense se ha visto afectada por una alta inestabilidad histórica y rigideces institucionales, marcada por conflictos de interés y cambios condicionados a los gobiernos de turno.

Además, la implementación de instrumentos de medición presenta debilidades técnicas y dificultades para articularse con la política curricular. Finalmente, su operacionalización no ha podido incluir las prácticas internacionales efectivas en la materia, lo cual ha limitado su capacidad de innovación. Todos estos vacíos han contribuido a que se impida el reconocimiento de la macroevaluación como un bien social de alto valor público que fomenta la mejora educativa y reduce las desigualdades. Los hallazgos se detallan a continuación.

Sistema reducido a la medición de los aprendizajes estudiantiles con escaso uso de sus resultados para la mejora educativa

En la sección anterior se evidenció que los componentes clave de un sistema de macroevaluación deben funcionar de manera interrelacionada para fomentar transformaciones educativas significativas y mejorar el logro de los aprendizajes estudiantiles. No obstante, aunque la macroevaluación ha sido parte de la agenda educativa del país a lo largo de su historia, el sistema se ha reducido a la medición del logro de los estudiantes y, además, presenta debilidades técnicas

que no han permitido la innovación educativa ni la incorporación de otros actores claves del sistema, como los centros educativos y el personal docente. Además, ha presentado deficiencias importantes para al difundir y transformar la información hacia una herramienta de gestión que promueva la mejora educativa.

Esta situación resulta crítica en un contexto como el actual, en el que se requiere con urgencia información de calidad para tomar decisiones específicas de política educativa que puedan contrarrestar los efectos producidos por el apagón educativo, así como mejorar los bajos desempeños que registra históricamente el país en competencias claves (como la lectora, matemática y científica) y reducir las desigualdades educativas.

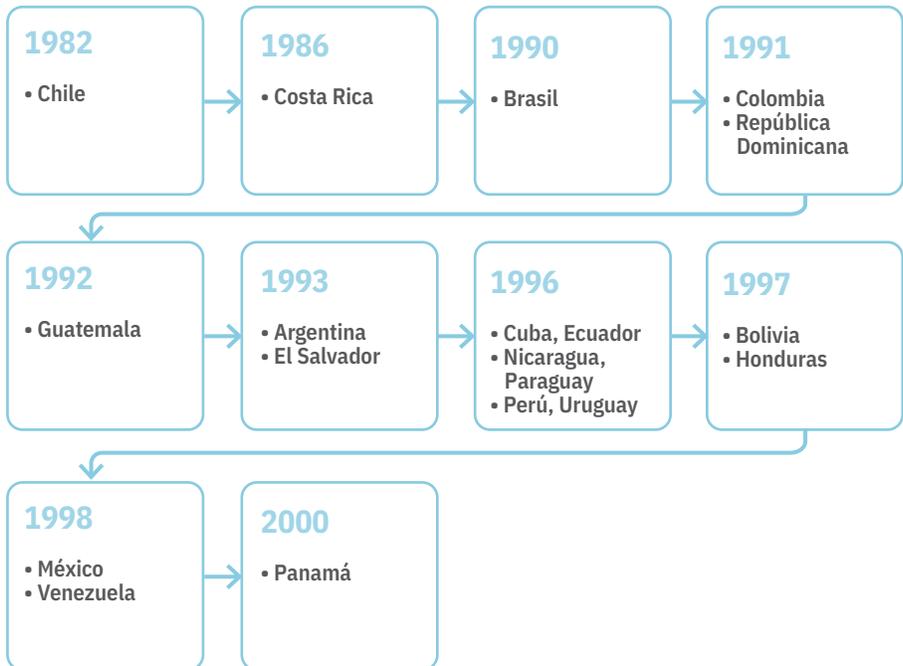
La macroevaluación en Costa Rica se reduce a un sistema medición del logro de aprendizajes

A pesar de que Costa Rica fue uno de los países de la región que reconoció tempranamente la importancia de la macroevaluación como factor relevante para acompañar los procesos de aprendizaje, solo superado por Chile (figura 3.5), cuando se analiza su trayectoria en este campo, salta a la vista que la experiencia se ha centrado casi exclusivamente en practicar mediciones estandarizadas del logro educativo de los estudiantes y a la fecha no ha logrado evolucionar hacia un sistema más comprensivo que funcione como herramienta de gestión para la mejora educativa.

Como se muestra en los siguientes párrafos, históricamente, la macroevaluación educativa en Costa Rica ha limitado su conceptualización y operacionalización al diseño y aplicación de pruebas estandarizadas con alcance nacional e internacional. Sin embargo, a pesar de que dicha aplicación es fundamental para dinamizar el sistema y ha demostrado ser de gran utilidad en el monitoreo del desempeño del sistema educativo y en el establecimiento de comparaciones con otros países, su información ha probado ser insuficiente para brindar soluciones y definir estrategias de mejora que permitan mejorar los desempeños, y com-

Figura 3.5

Inicio de aplicaciones de evaluaciones estandarizadas en América Latina



Fuente: Barquero, 2023 a partir de Galas et al., 2020.

batir la desigualdad o brechas educativas (Unesco, 2008). Por su parte, la institucionalidad a cargo de la macroevaluación nacional se centró únicamente en el MEP, de manera que la responsabilidad de ejecutar las políticas educativas y de evaluarlas recayó en un mismo ente y, en principio, en un mismo jerarca, lo cual representa una problemática que se detallará más adelante.

Un trabajo realizado por Francis (2023a) muestra que el sistema nacional de macroevaluación educativa costarricense se ha desarrollado en cinco etapas (figura 3.6). Los inicios de la macroevaluación en Costa Rica se remontan a un largo período que abarca finales del siglo XIX y primeros dos tercios del siglo XX y estuvieron a cargo de los institutos nacionales de segunda enseñanza y estudios normalistas. Se trató de experiencias puntuales con dos propósitos específicos: permitir la admisión a la segunda enseñanza y proveer la certificación del título de Bachiller en Ciencias o en Letras y de Maestro Normal. Estas pruebas tenían

un enfoque sumativo cuyo alcance no era promover el mejoramiento del sistema educativo, sino filtrar a las personas capacitadas para cursar la enseñanza superior e incorporarse para mejorar la calidad política y técnica del país. Es decir, sus resultados no se utilizaron como insumos para la evaluación del sistema ni para la toma de decisiones en las distintas reformas que se impulsaron en este periodo.

Una segunda etapa histórica transcurrió entre 1974 y 1985, cuando se eliminaron las Pruebas de Bachillerato y, por tanto, las acciones evaluativas se centraron en el ámbito de las instituciones educativas específicas. Aunque el Consejo Superior de Educación (CSE) y el Ministerio de Educación Pública (MEP) fueron definidos por la normativa como los responsables del mejoramiento cualitativo de la educación, sus políticas y decisiones no tomaron en cuenta los resultados de las evaluaciones realizadas en la fase anterior. La decisión de eliminar las pruebas buscaba ser coherente con

Figura 3.6

Fases del desarrollo del Sistema de Macroevaluación Educativa en Costa Rica



Fuente: Francis, 2023a.

las tendencias educativas de la época, las cuales se enfatizaban el desarrollo de las personas y la oportunidad de contribuir con una formación permanente, pero no otorgaban relevancia a la macroevaluación como herramienta para lograr esos fines. Este período fue seguido por una corta etapa —la tercera—, en la que se reinstauraron las pruebas nacionales y se establecieron las bases del sistema actual de evaluación.

La cuarta etapa del desarrollo histórico de la evaluación educativa se dio entre 1988 y 2018. A partir de 1988, se adoptaron diversas medidas orientadas hacia el desarrollo de un sistema de evaluación nacional que fungiera como herramienta clave para el aprendizaje de los estudiantes (OCDE, 2017a). Esta decisión se fundamentó en los bajos desempeños estudiantiles que arrojaban los estudios diagnósticos realizados por entidades académicas de las universidades y por la carencia de mecanismos en el MEP para garantizar el uso y cumplimiento de los programas estudio vigentes en ese momento Francis (2023a).

Se creó el sistema de información escolar conocido como el Sistema Nacional de Evaluación de la Calidad de la Educación (SNECE) y el Modelo de Evaluación de la Calidad de la Educación Costarricense (Mecec) para monitorear la calidad de los procesos educativos en las instituciones y la Dirección de Evaluación y Gestión de Calidad (DEGC) al interior del Ministerio de Educación Pública (MEP). También, se reinstaló la aplicación de pruebas nacionales con carácter diagnóstico y de certificación en primaria y secundaria, respectivamente, de acuerdo con las políticas y normativas vigentes. Además, a partir del 2009, el país inició su participación en las pruebas estandarizadas internacionales del Programa Internacional para la Evaluación de Estudiantes (PISA) de la OCDE, un paso inédito y relevante que le permitió al país comenzar a generar información comparada con los mejores sistemas educativos del mundo.

Durante este período, el propósito de la macroevaluación se orientó hacia el empleo de los resultados para fomentar

la transparencia, la rendición de cuentas y el mejoramiento educativo de manera agregada para el sistema educativo en su conjunto. La evaluación también se promovió como un mecanismo para motivar a los alumnos a aumentar sus esfuerzos y sus logros. No obstante, los esfuerzos técnicos se concentraron en el modelo de medición y no en el análisis y uso de los resultados.

Actualmente, el país atraviesa una quinta etapa en el desarrollo histórico de la macroevaluación. Esta inició en 2019 y se caracteriza por dos aspectos importantes: la suspensión temporal de las mediciones nacionales estandarizadas y la reconceptualización de un nuevo modelo de macroevaluación educativa que no llegó a consolidarse. Se trata de una etapa de fuerte incertidumbre y zigzagueo desde el punto de vista de la política educativa.

En cuanto a la suspensión, destacan dos momentos importantes marcados por decisiones políticas de los gobiernos de turno. El primero ocurrió entre 2019-2020, cuando la Administración Alvarado Quesada decide eliminar las Pruebas Nacionales de Bachillerato vigentes desde 1988 y sustituirlas por las Pruebas FARO; y el segundo, en 2022, bajo la Administración Chaves Robles, cuando se eliminaron las recién inauguradas pruebas FARO y se sustituyeron por la Prueba Nacional Estandarizada en 2023, cuya implementación se encuentra actualmente en curso.

La Administración Alvarado Quesada incorporó, en cuanto a la reconceptualización metodológica, la aplicación de pruebas estandarizadas progresivas⁴ en los últimos años de la primaria y secundaria, la implementación de cuestionarios de factores asociados al rendimiento académico, propuestas de evaluación de desempeño docente y la generación de estándares educativos como base para el andamiaje de pruebas nacionales y la evaluación de los aprendizajes en el aula. Pese a que estas medidas eran disruptivas e iban en concordancia con las mejores prácticas en macroevaluación (Ospina y Ángulo, 2023) las debilidades y fuertes cuestionamientos técnicos realizados a algunos de los instrumentos afectaron su consolidación.

Según el último informe de la Contraloría General de la República, el Sistema Nacional de Evaluación de la Calidad de la Educación vigente no cumple con el marco regulatorio debido, entre otros factores, a la ausencia de una visión estratégica y a debilidades en la implementación del Modelo de Evaluación de la Calidad de la Educación. Señala, además, que actualmente no se dispone de información útil, confiable y oportuna para rendir cuentas, mejorar continuamente la prestación del servicio educativo y tomar decisiones en cuanto a la distribución de recursos, la atención a la población vulnerable y la priorización de estrategias para reducir las desigualdades educativas (CGR, 2022).

Escasa difusión y uso de resultados: un vacío histórico

El sistema de macroevaluación costarricense no solo se ha reducido a la medición de los aprendizajes, sino que también ha carecido de capacidades administrativas y técnicas que le permitan transformar los resultados obtenidos en información y conocimiento para la mejora de la calidad y equidad educativa. De acuerdo con la literatura, la influencia que ejerzan las evaluaciones de los aprendizajes en la calidad de la educación depende fundamentalmente del modo en que se presentan los resultados, la información que se comparte y de que los hallazgos sean utilizados por distintas audiencias (escuelas, docentes, administrativos, padres de familia y la sociedad civil en general) (Ferrer y Fiszbén, 2015). En esta línea, Lasida (2023) señala que comunicar los resultados le da sentido a la evaluación en tanto posibilita el logro de sus dos principales propósitos: la mejora de la educación y la rendición de cuentas en todos los niveles, como derecho y responsabilidad de todos los actores⁵ vinculados a la educación.

El análisis realizado por Francis (2023a) permite concluir que el uso de las evaluaciones en el país se ha concentrado en concebirlas como herramientas para la gestión (para más información sobre la tipología de usos de la macroevaluación ver recuadro 3.2). Dicha función ha estado inscrita desde la implementa-

ción de las pruebas nacionales en 1988. Sin embargo, el alcance ha sido bajo debido a que los esfuerzos se han limitado a informar a la opinión pública sobre el estado de la educación a nivel nacional o internacional, con un escaso impacto en las decisiones de política educativa que se han basado principalmente en modificaciones específicas a los temarios de evaluación o al diseño de las pruebas (Francis, 2023a y Ferrer, 2021, 2006; González et al., 2018, Guzmán et al., 2006).

La normativa vigente⁶ en los procesos de macroevaluación señala en el artículo 88 la responsabilidad de la DGEC de remitir a los centros educativos el informe de resultados de las pruebas nacionales. El propósito de tal informe es servir como insumo para los distintos actores educativos en la mejora del proceso educativo, así como para las autoridades en la toma de decisiones que coadyuven en el seguimiento de estrategias de mejora continua del proceso de educativo y de la mediación pedagógica.

Mayoritariamente, los usos y difusión se han restringido a la elaboración de informes nacionales que describen la evaluación y de las competencias que se miden; además, se ofrecen estadísticas descriptivas de los examinandos, resultados promedio, así como comparaciones entre grupos poblacionales o direcciones regionales educativas (Francis, 2023a). No obstante, Ospina y Ángulo (2023) señalan que, a pesar de que estos informes son rigurosos, pueden no ser muy útiles para los diferentes tomadores de decisiones, pues se espera que sean los diferentes actores quienes realicen la interpretación de los resultados de acuerdo con sus propios intereses.

En relación con lo anterior, un estudio realizado por Francis (2023b) para este Informe, recopiló las voces⁷ de los actores educativos (personal docente, directores, directoras y personas supervisoras) con el fin de comprender la importancia que le otorgan a la información derivada de la macroevaluación y cómo esta debería ser presentada para que sea útil para el personal docente. Este último aspecto se aborda con detalle en la última sección de este capítulo.

Recuadro 3.2

Tipología de usos de la información derivada de los procesos macroevaluativos

- **Como herramienta para la gestión:** se concibe como un componente de aseguramiento de la calidad a partir de la cual los actores educativos buscan mejorar el servicio que prestan.
- **Como objeto de análisis psicométrico:** los esfuerzos se concentran en producir información técnicamente sólida, válida y útil para los tomadores de decisión; se asume que estos tienen preguntas que necesitan responderse y que coinciden con las evaluaciones desarrolladas.
- **Como fuente de información pedagógica:** el interés está en relacionar la evaluación con lo que hacen tanto docentes como estudiantes.

Fuente: Fortín, 2013.

Los principales hallazgos se enumeran a continuación. En primer lugar, los actores educativos consideran que la macroevaluación genera cantidades significativas de información y datos que potencian la mejora de las prácticas en todo el sistema educativo. Para lograrlo, los informes deben dirigirse a todo el personal docente, no solo a quienes están cargo de los niveles en los que se aplican las pruebas. Además, deben plantearse como mecanismos estratégicos de mejora didáctica y no solo como un informe de situación, tal y como sucede en la actualidad. En tercer lugar, señalan que la gobernanza debe reconocer a los y las docentes como audiencia clave en los procesos macroevaluativos. Esto implica considerarlos como actores centrales o destinatarios de las decisiones y acciones derivadas de dichos procesos. Finalmente, se destaca la importancia de generar, con base en la lectura de los resultados de la macroevaluación, una

mayor autonomía en los centros educativos para impulsar el desarrollo institucional, los planes de mejora y los proyectos de innovación, entre otros.

Lo anterior concuerda con lo señalado por otros estudios. Por ejemplo, Clarke y Luna (2021) sugieren que, para satisfacer las necesidades de distintos actores interesados, puede ser necesario elaborar varios productos de comunicación en lugar de un único informe. Esto se debe a la diversidad de intereses y capacidades técnicas entre los grupos interesados, lo que puede presentar un desafío ante la difusión de los resultados de la evaluación nacional a gran escala. Además, Greaney y Kellaghan (2008) destacan la importancia de comunicar eficazmente los resultados de la evaluación al público en general para influir en la opinión pública en cuanto a los cambios propuestos en la política educativa y en dar a conocer el estado actual de las prácticas educativas.

A pesar de estas recomendaciones, en el país la macroevaluación ha sido entendida erróneamente como un fin en sí misma, lo cual ha ocasionado que toda la dinámica formativa en el aula se ajustara a lo que las pruebas midieran. En consecuencia, se perdieron de vista las intencionalidades curriculares previstas en los programas de estudio (Guzmán et al., 2006). Estas implicaciones no resultan coherentes con los propósitos establecidos en la política curricular, los cuales apuntan hacia un uso sistemático de la información generada para sustentar la toma de decisiones así como brindar insumos para su actualización.

La situación descrita se complica al considerar los cuestionamientos técnicos que han surgido en torno al sistema de medición de los logros de aprendizaje en el país. Estos cuestionamientos se detallarán en las próximas secciones de este Informe. Por el momento, puede señalarse que la calidad técnica de estos procesos se relaciona estrechamente con la difusión y el uso de los resultados generados en la macroevaluación (cuadro 3.1). Una deficiente calidad técnica conduce a efectos perversos en el uso de la información, pues será la base para tomar decisiones imprecisas y emplear de forma ineficiente los recursos, tal y como

Cuadro 3.1

Relación entre la calidad técnica y la difusión y usos de los resultados en la macroevaluación

Difusión y uso	Calidad técnica	
	Buena	Deficiente
Adecuados a las características técnicas de la prueba	Aporte. Posibilidad de impacto en la mejora del rendimiento en las áreas evaluadas	Efectos perversos, interpretaciones y decisiones erradas
No adecuados a las características técnicas de la prueba	Efectos perversos, interpretaciones y decisiones erradas	Efectos perversos, interpretaciones y decisiones erradas
Escasos o nulos	Desperdicio de recursos y oportunidades	Desperdicio de recursos

Fuente: Galas et al., 2020.

se expone en el cuadro 3.1.

Desde esta perspectiva, el sistema educativo costarricense debe prestar atención urgente a la difusión y uso de los resultados de la macroevaluación. Para lograrlo, debe contar con una incuestionable capacidad técnica que le permita brindar información acorde a las necesidades de quienes la reciben. De esta forma, los diversos actores de la educación tendrán bases sólidas para comprender e incidir en la transformación educativa. En este sentido, Fortín (2013) indica que los principales desafíos por atender tienen que ver con garantizar que la información llegue a los distintos usuarios, que se represente de forma clara y accesible para facilitar la comprensión entre todos los actores y que se proporcione con el claro propósito de incentivar cambios a favor de la calidad educativa. De acuerdo con Galas et al., (2020) no solo se debe trabajar en mejorar y diversificar los formatos y el alcance en cómo se comunican los resultados, sino que también es necesario atender la calidad técnica y comunicativa para el público al que se dirige. Es decir, se deben realizar esfuerzos por educar a las audiencias usuarias de los resultados de las evaluaciones para facilitar su comprensión y uso pedagógico.

En el caso de Costa Rica, ambas funciones, la difusión y el uso de resultados, están actualmente a cargo de

instancias distintas: la primera función le corresponde a la DGEC y la segunda a la Dirección de Desarrollo Curricular. Estas entidades han operado con poca articulación, lo que ha promovido la aparición de rupturas u obstáculos en la continuidad del proceso evaluativo (Rodino, 2023a). De no atender la desconexión entre el diseño de las evaluaciones, su implementación, la adecuada difusión de los resultados, el manejo contextual de las fuentes y el uso pertinente de los datos se deberá hacer frente a graves consecuencias en el sistema educativo, entre las que se encuentran la generación de desconfianza en el proceso y la puesta en entredicho de su importancia para impactar en la mejora de los aprendizajes Ospina y Ángulo (2023).

Para más información sobre

REQUIRIMIENTOS DE LOS INFORMES DE LAS PRUEBAS NACIONALES PARA RETROALIMENTAR LA MEDIACIÓN DOCENTE

véase Francis, 2023b

www.estadonacion.or.cr

Limitaciones en el marco teórico y normativo reducen el alcance de la macroevaluación

Para implementar un enfoque sistémico en la macroevaluación de los aprendizajes, resulta fundamental contar con un marco teórico y normativo adecuado, ya que este determina el alcance del sistema y condiciona su funcionamiento. Pese a dicha necesidad, el marco de referencia actual posee vacíos importantes en sus componentes, conceptualización, participación de actores y articulación entre procesos sumativos y formativos.

En primera instancia, el marco actual se enfoca únicamente en la evaluación de estudiantes mediante pruebas nacionales, lo cual deja de lado otros componentes como la evaluación docente y la del centro educativo. En segundo lugar, se han detectado limitaciones para definir, oficializar y divulgar el propósito estratégico, los objetivos y metas que fundamentan el diseño de una estrategia de evaluación de calidad con una visión sistémica hasta ahora ausente en su funcionamiento (CGR, 2022).

Por otra parte, según Ospina y Angulo (2023), el marco actual no precisa cuál debería ser la participación de los diferentes actores de la comunidad educativa en la implementación y el análisis de los resultados. Tampoco existe claridad en cuanto a la gobernanza y los objetivos del sistema de evaluación, ni tampoco en el cronograma de tiempos y requisitos

necesarios para garantizar la continuidad, la coherencia con los objetivos, la comparabilidad entre los datos y su análisis y la apropiación por parte de la comunidad educativa.

Además, señalan vacíos en la propuesta en cuanto a la información y herramientas necesarias para articular la evaluación diagnóstico-formativa y la sumativa. A pesar de que estas funciones están expresadas en el Reglamento de Evaluación de Aprendizajes vigente, la formulación de las pruebas mantiene la esencia de las normativas vigentes desde 1892, pues se interpretan como un proceso de evaluación censal y de carácter sumativo para acreditar o seleccionar a alumnos de un curso determinado.

En Costa Rica, tanto la conceptualización como los propósitos han evolucionado a lo largo de su historia (cuadro 3.2). Es posible notar como la definición y objetivos principales del sistema han pasado de centrarse en la evaluación y certificación de los aprendizajes hacia una concepción de control de calidad y para fundamentar la toma de decisiones hacia el mejoramiento educativo.

Los diversos marcos normativos han estado determinados por diferentes reglamentos de evaluación (figura 3.7) y han evolucionado en función de los objetivos previamente indicados en el cuadro 3.2. Estos reglamentos han estado bajo la responsabilidad del MEP y han regulado la macroevaluación de aprendizajes en el

país, a excepción de los reglamentos aprobados en 1974 y 1976, los cuales fueron implementados en una época sin evaluaciones estandarizadas, como se mencionó anteriormente.

Antes del 2004, se crearon seis reglamentos y no fue sino hasta después de ese año cuando se consolidó el Reglamento de Evaluación de Aprendizajes, vigente en la actualidad como parte de la política educativa. A partir del 2019, el Reglamento experimentó cambios a raíz de los nuevos propósitos y objetivos que fueron anteriormente expuestos, pero conservó la naturaleza y alcance de las pruebas nacionales, las cuales hasta ese momento habían gozado de una relativa estabilidad en su implementación durante las dos décadas previas. La puesta en marcha de estos cambios se atrasó hasta 2021 debido a los efectos de la huelga del año 2018, un conflicto social que tuvo una duración de 93 días (MEP, 2019), y de la crisis ocasionada por la covid-19 que implicó a la redefinición de acciones educativas, entre ellas la macroevaluación educativa (Francis, 2023a).

El reglamento vigente define la evaluación de aprendizajes como “un proceso continuo de recopilación de información cualitativa y cuantitativa, que fundamenta la emisión de juicios de valor y la toma de decisiones por parte de la persona docente y el estudiantado, para la mejora progresiva de los procesos de enseñanza y aprendizaje” (MEP, 2020). Asimismo,

Cuadro 3.2

Conceptualización y objetivos del Sistema de Macroevaluación de Costa Rica. 1892-2022

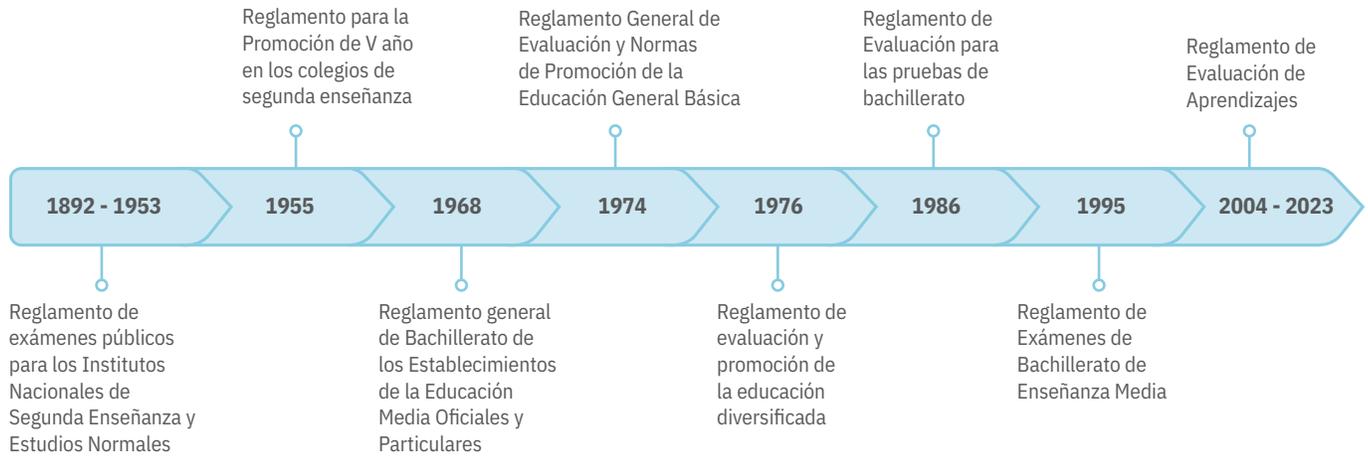
Etapa	Conceptos	Objetivos
1892 -1973	Evaluación como medición de saberes académicos.	Certificar la culminación de la educación media.
1974-1985 ^{a/}	Evaluación de aprendizajes de nivel o grado.	Certificar los aprendizajes logrados para la promoción de grado o nivel.
1986-2018	Evaluación proceso de naturaleza sistemático que coadyuva en la valoración de la calidad del sistema educativo.	Control de calidad y acreditación de logros básicos. Establecer, readecuar y valorar la calidad de la educación.
2019- 2022	Proceso continuo de recopilación de información cualitativa y cuantitativa, que fundamenta la emisión de juicios de valor y la toma de decisiones.	Mejoramiento progresivo de los procesos de enseñanza y aprendizaje y certificación.

a/ En esta etapa, el sistema educativo no aplicó pruebas estandarizadas: la toma de decisiones se basó en las estadísticas institucionales recolectadas a la fecha y con base en el Plan Nacional de Desarrollo vigente en cada gobierno.

Fuente: Barquero, 2023 con base en Francis, 2023a.

Figura 3.7

Reglamentos de evaluación de aprendizajes aprobados en el sistema educativo costarricense. 1892-2023



Fuente: Barquero, 2023 con base en Francis, 2023a y MEP, 1973.

estipula que sus funciones en el proceso educativo son de carácter diagnóstico, formativo y sumativo. La función diagnóstica permite conocer el estado inicial de la persona estudiante en las áreas del desarrollo cognoscitiva, socioafectiva y psicomotriz, con el fin de facilitar, a partir de la información que de ella se deriva la aplicación de las estrategias correspondientes. La función formativa brinda información necesaria y oportuna durante los procesos de enseñanza y aprendizaje con la finalidad de reorientar o realimentar las áreas que así lo requieran. Finalmente, la función sumativa constata los logros alcanzados al término de un proceso de aprendizaje y fundamenta la calificación, la promoción y la certificación (MEP, 2023). Esta última, según lo indica el Reglamento, se vincula a los objetivos de las pruebas nacionales definidos en su artículo 81^o. Estas claras definiciones conceptuales no han sido suficientes, sin embargo, para consolidar mecanismos que integren los resultados y que se gesten los cambios requeridos, dado que el MEP no ha podido concretar estrategias de uso de dicha información (Francis, 2023a).

Por tanto, se requiere mejorar los aspectos señalados que limitan el enfoque sistémico bajo el que debería operar la macroevaluación. Si bien esto implica

la construcción de un sistema más complejo (por las implicaciones en los diseños de instrumentos los cuales deben ser distintos para responder efectivamente a cada propósito), a largo plazo favorecerá a contar con más y mejor información que permita evaluar y asegurar la calidad de los aprendizajes de los estudiantes, mediante la realización de los ajustes necesarios tanto a nivel de sistema, como a nivel de las salas de clases y centros educativos (Medeiros, 2023).

Otro vacío que detectado en el marco de referencia actual tiene que ver con que este no incluye la generación de estándares educativos, como sí los hay en otros países. Dichos estándares clarifican los logros de aprendizaje que se pretenden alcanzar y que se recomienda que sean la base del currículo (Lasida, 2023). Aunque el país había empezado a trabajar en este asunto (recuadro 3.3), su ausencia es un vacío importante en el sistema actual y retomarlo es tanto una responsabilidad como un desafío de las autoridades educativas.

Inestabilidad y rigideces institucionales obstaculizan la consolidación del sistema

La trayectoria de la institucionalidad a cargo de la macroevaluación educativa

en el país ha estado marcada por dos aspectos importantes. En primer lugar, el MEP ha liderado el funcionamiento del sistema, pero ha existido una alta inestabilidad en cuanto a la institución o dependencia específica encargada de sus funciones. En segundo lugar, esta situación ha generado un conflicto de interés en la política pública, ya que la instancia que ejecuta esta política es también la encargada de la evaluación de sus resultados. Como consecuencia, las personas evaluadoras están subordinadas al funcionario responsable de la ejecución (es decir, el MEP), lo cual genera una situación compleja en términos de imparcialidad y objetividad en la evaluación (Ospina y Angulo, 2023).

Como se mostrará a continuación, hubo consecuencias negativas que obstaculizaron el progreso del sistema, tales como la falta de continuidad orgánica de las políticas, los cuestionamientos a la gobernanza del sistema, el aprovechamiento de la información por parte de la comunidad educativa, los cambios constantes en las evaluaciones e incertidumbre en su permanencia y dificultades para innovar e implementar transformaciones hacia la mejora de los aprendizajes.

Recuadro 3.3

Importancia de los estándares educativos para evaluar el progreso de la educación

Un requisito básico para el mejoramiento educativo es establecer las metas de aprendizaje específicas que se deben alcanzar. Los estándares educativos de aprendizaje son descripciones escritas y concisas de lo que se espera que el estudiantado conozca y sepa hacer en cada etapa específica de su proceso formativo. Describen objetivos educativos como, por ejemplo, lo que los y las estudiantes deberían haber aprendido hacia el final del curso, el grado o el período lectivo, aunque, en general, no están asociados a una práctica particular de enseñanza (Hambleton, 2001; Bomer y Maloch, 2011; Griffin, McGaw y Care, 2012; Common Core State Standards Initiative, 2010). Además, los estándares deben estar alienados con los programas de estudio y considerar todos los aprendizajes esperados que se establecen en ellos. Deben ser comprensibles para las diversas audiencias de la educación, de tal forma que toda la sociedad pueda estar al tanto de qué es lo que se busca en términos de aprendizaje en cada una de las asignaturas y niveles de enseñanza.

Las pruebas, tanto de aula como estandarizadas, deben construirse partiendo de los estándares educativos para poder obtener inferencias precisas sobre los niveles de competencia de las personas examinadas y no solamente generar un valor numérico asociado al desempeño en el examen. Estas descripciones deben compartirse con estudiantes y sus familias, además del personal docente y las autoridades educativas, para que todas las instancias involucradas en el proceso sean participantes activos del proceso (Common Core State Standards Initiative, 2010; National Research Council, 2001; Ravela et al., 2008).

Las características deseables en un estándar educativo (Perie, 2018; Schneider y Egan, 2013) son:

- Expresar una capacidad o competencia, un saber hacer, no solamente listar un conocimiento.
- Describir un desempeño concreto que se espera observar en el estudiante.
- Estar redactado de manera clara y precisa, de manera que pueda ser entendido de manera homogénea por las diversas audiencias.
- Ser unidimensional, es decir medir un solo rasgo o atributo.
- Proveer indicios claros en torno a su evaluación.
- Referirse a un aprendizaje sustantivo, no a una acción o actividad.
- Evitar términos subjetivos como: “satisfactorio”, “bueno”, “aceptable”, “adecuado”, “menos”, “más” o “mejor”, pues pueden significar diferentes cosas para diferentes personas.

Aunque, históricamente, lo más común ha sido establecer estándares educativos de carácter cognitivo en términos de conocimientos, destrezas o habilidades, también es posible construir estándares de naturaleza afectiva, siempre que existan metas de aprendizaje en dimensiones socioemocionales y actitudinales (López, 2011; Manning et al., 2017, Arias et al., 2020). Los estándares cognitivos contestan a preguntas como la siguiente: ¿qué deben saber hacer los estudiantes al finalizar cada periodo o año lectivo?, mientras que los afectivos responden a este tipo de interrogantes: ¿qué actitudes y habilidades socioemocionales deben haber desarrollado los estudiantes al finalizar cada año o periodo lectivo?

La experiencia internacional y nacional

La comunidad de medición educativa de los Estados Unidos, a través de su asociación

profesional NCME (*National Council on Measurement in Education*, <https://www.ncme.org/home>) formalmente inició la discusión técnico-científica en torno a la necesidad de estándares educativos explícitos como una vía para solventar el divorcio que ya en esa época se notaba entre evaluación estandarizada y de aula⁹. A partir de ese momento, el tema se dio a conocer internacionalmente. A raíz de lo anterior, en Estados Unidos se desarrolló la iniciativa denominada *Common Core Standards*, que se reconoce actualmente como uno de los temas más importantes en la educación pública de ese país y que influye en cada dimensión de sistema educativo: las pruebas estandarizadas de alto impacto, las destrezas que se enseñan a los estudiantes en los centros educativos y el desarrollo profesional del cuerpo docente (Peterson et al., 2017).

A nivel internacional, una cantidad importante de países han optado por su establecimiento. Entre ellos, en América Latina, se encuentran Colombia, Ecuador, Honduras, Guatemala, Chile y Uruguay. Estos dos últimos son los países latinoamericanos con mejores resultados en las pruebas PISA 2018.

En Costa Rica se han realizado dos importantes esfuerzos para avanzar en su elaboración. El primero tuvo lugar a finales de los años 90 del siglo XX, cuando se promovió el *Proyecto de establecimiento de Estándares para la Educación Primaria en Centroamérica, s.f.*, para establecer estándares de contenido y desempeño en cada año de primaria y secundaria en el marco del SEC-SICA, a nivel centroamericano. Sin embargo, este proyecto no llegó a implementarse. El segundo esfuerzo se desarrolló entre el 2020 y parte del 2022 en el MEP, con el proyecto Estándares Educativos y Niveles de Desempeño del Ministerio de

Recuadro 3.3 (continuación)

Importancia de los estándares educativos para evaluar el progreso de la educación

Educación Pública de Costa Rica promovido con la participación la Dirección de Desarrollo Curricular (DDC) y a la Dirección de Gestión y Evaluación de la Calidad (DGEC) como una necesidad de alinear las pruebas estandarizadas con la transformación curricular.

Este proyecto se basó en un enfoque sistémico, pues la DDC, como encargada de elaborar e implementar los programas de estudio, y la DGEC, como responsable de construir y aplicar las pruebas estandarizadas nacionales, habían venido trabajando de forma relativamente autónoma en sus propuestas. Es decir, existía una disonancia importante entre las propuestas para la mediación y la evaluación en el aula y la evaluación por medio de pruebas estandarizadas¹⁰ (Montero, 2020).

La propuesta de estándares educativos generada por este proyecto tuvo como base el análisis y la operacionalización de los programas de estudio, la Política Curricular vigente y otros documentos oficiales apro-

bados por el Consejo Superior de Educación y brindan la descripción del desempeño educativo esperado del estudiantado al concluir cada año lectivo. Su objetivo fue contar con marco de referencia común, tanto en la mediación y la evaluación de aula como en la evaluación por medio de pruebas estandarizadas. También, se construyeron con el objetivo de que fueran comprensibles para todos los actores relevantes, incluyendo estudiantes, padres y madres de familia.

La propuesta incluye niveles de desempeño que acompañan a cada estándar, los cuales sugieren una propuesta básica de trayectoria para lograr el aprendizaje, lo que provee a los docentes de orientaciones pedagógicas para su trabajo en el aula. Los niveles de desempeño también resultan útiles para realizar evaluaciones diagnósticas. Para la creación de los estándares, se realizó un análisis y operacionalización los documentos mencionados arriba con el fin de que fueran observables y evaluables, cubrieran a cabalidad los programas de estudio y permitieran valorar el rendimiento o logro alcanzado por medio de

los niveles de desempeño. También se elaboraron perfiles generales de desempeño, los cuales consisten en expectativas de aprendizaje de naturaleza más resumida y global para la difusión hacia el público general y los diversos actores no especializados del sistema educativo.

La propuesta abarcó desde primer grado a undécimo año de la modalidad académica diurna en Ciencias, Español y Matemática y fue sometida a valoración externa y validada por expertos de las universidades. Las actas del Consejo Superior de Educación No. 28-21 del 17 de mayo y 41-21 del 12 de agosto de 2021 documentan este proceso. Una vez concluida la valoración externa, se presentará de nuevo para su aprobación final por parte del CSE. Sin embargo, a la fecha de realización del informe dicho proceso no ha mostrado ningún avance.

Fuente: Montero, 2023.

Múltiples cambios institucionales afectan la gobernanza de la macroevaluación

El proceso evaluativo está influenciado por factores históricos y culturales que lo moldean y, por lo tanto, es necesario examinar las tensiones, los problemas y las oportunidades que surgen de esta conexión para poder otorgar un papel estratégico a sus contribuciones (Francis, 2023a). Como se mencionó, en Costa Rica este proceso lo ha liderado por el MEP, pero ha experimentado cambios recurrentes en la institucionalidad o instancia específica a cargo dentro del Ministerio. Estos cambios han afectado principalmente las funciones del sistema, su gobernanza, el aprovechamiento de su información y su capacidad de innovación (Ospina y Angulo, 2023; Francis, 2023a).

Este hallazgo resulta de interés particular porque la institucionalidad condiciona todos los componentes del sistema. Su estructura define no solo los liderazgos políticos sino además las responsabilidades políticas y administrativas en el diseño, ejecución y usos de los programas de evaluación estandarizada. Además, afecta la función de evaluación y sus efectos inciden en la credibilidad y la legitimidad de las actividades de evaluación, así como en la sostenibilidad técnica y operativa de estas. También, la institucionalidad determina la forma y el alcance de la información de los datos evaluativos, afecta la estabilidad de los equipos técnicos a cargo del diseño y aplicación de pruebas, la asignación y manejo de recursos presupuestados, los canales de comunicación entre evaluadores y usuarios de la información, el manejo

político en torno a las evaluaciones y los mecanismos de acceso público por parte de las audiencias.

Las funciones de la gobernanza en la evaluación se han concentrado más en los procesos de logística de la implementación de pruebas y en la elaboración de informes que en la promoción del uso de los resultados para la toma de decisiones (cuadro 3.3). Este hecho, además, produjo un impacto negativo para el desarrollo del sistema ante la ausencia de consenso entre los actores educativos, los cuestionamientos a su credibilidad y los cambios constantes en las evaluaciones, condicionados por las decisiones de los gobiernos de turno.

Francis (2023a) encuentra que el MEP ha desempeñado un protagonismo histórico en la administración y gestión de la macroevaluación (figura 3.8), pero

Cuadro 3.3

Funciones de la gobernanza según la etapa del desarrollo de la macroevaluación educativa en Costa Rica

Etapa	Funciones
Etapa 1: 1892-1973	Definir contenidos, configurar contenidos de pruebas, definición de Tribunales o Comisiones examinadoras, Calificación.
Etapa 2: 1974-1985	Programar los procesos de evaluación de aprendizajes.
Etapa 3: 1988-2018	Diseñar las pruebas con base en temarios aprobados por el CSE.
	Realizar la logística de aplicación.
	Elaborar informes.
	Analizar las pruebas internacionales.
Etapa 4: 2019 actualidad	Brindar asesoría a Centros educativos
	Diseñar las pruebas.
	Realizar la logística de aplicación.
	Elaborar informes de resultados.
	Analizar las pruebas internacionales.
	Asesorar a centros educativos con información de carácter administrativo y técnico para garantizar la correcta ejecución de las pruebas nacionales y el uso de la información.

Fuente: Barquero, 2023 con base en Francis, 2023a.

con cambios importantes en cuanto al rol de la institución o dependencia específica del Ministerio que estuvo a cargo. En la primera etapa de la macroevaluación (1892-1973), la responsabilidad de la preparación de las pruebas recaía en los Institutos Nacionales de Primera Enseñanza. Sin embargo, con la creación de nuevos reglamentos¹¹ en evaluación, esta responsabilidad se transfirió al MEP a través del Departamento de Supervisión General, el cual se encargaba de orientar, coordinar y evaluar las actividades educativas. A pesar del apoyo brindado por instancias como la Dirección Curricular y el Departamento de Supervisión General, se produjo una centralización de los procesos de preparación, edición, ensamblaje y empaque de las pruebas que no resultó en una mejora en la gestión. Ante esto, al finalizar el periodo, el CSE decidió que la preparación de las pruebas volviera a estar a cargo de los Institutos Nacionales de Primera Enseñanza, debido a cuestionamientos sobre la capacidad de la gobernanza para asumir estos procesos.

La segunda etapa abarcó una década en la que el sistema educativo no implementó procesos de macroevaluación educativa. Las decisiones de política educativa estuvieron vinculadas a las metas establecidas por los gobiernos de turno en sus respectivos planes nacionales de desarrollo. Se concentraron esfuerzos en la creación de estructuras orgánicas (la Unidad Estadística y la Unidad de Investigaciones Pedagógicas) y en la promulgación de reglamentos de evaluación y estudios con entes externos para generar información general del sistema con nula vinculación a los procesos de macroevaluación. Estas medidas no generaron las mejoras esperadas en materia de calidad y nuevamente las tendencias del país se mantenían en declive.

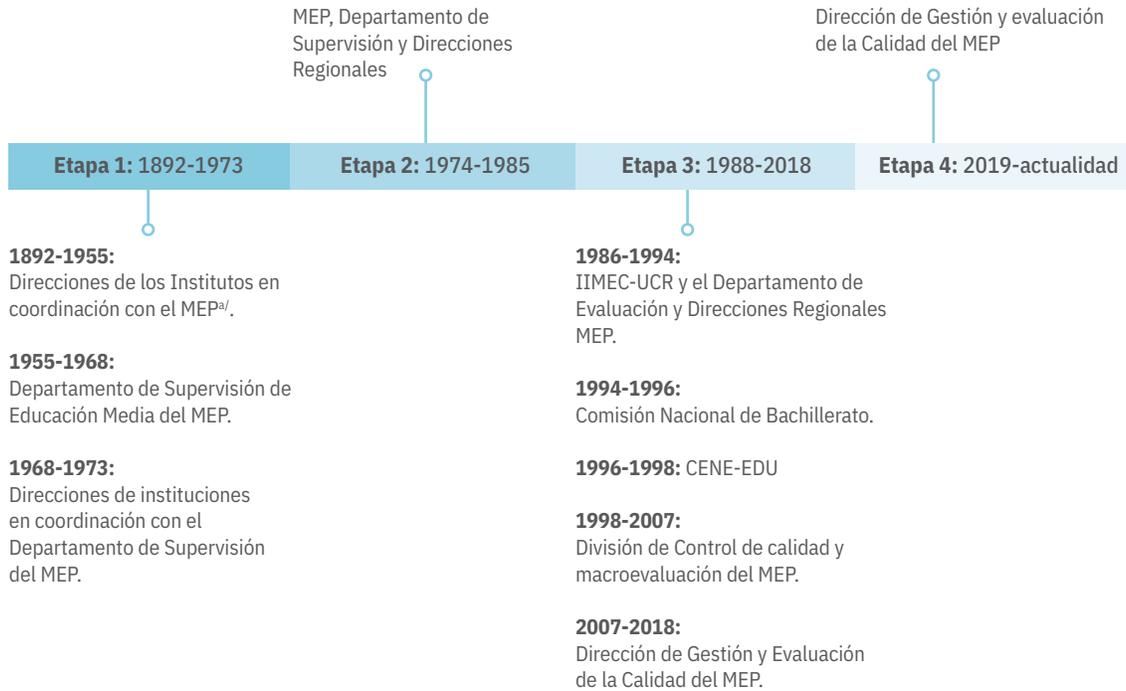
La tercera etapa se caracterizó por la restauración y el reconocimiento de la macroevaluación como un mecanismo de monitoreo, transparencia, rendición de cuentas y como un proceso generador de información para el mejoramiento de calidad educativa. Este reconocimiento estuvo acompañado por dos hechos

relevantes: en primer lugar, se llevaron a cabo una serie de reestructuraciones en la gobernanza con el fin de consolidar un sistema de gestión que permitiera la convergencia del MEP. En segundo lugar, se promovió el trabajo colaborativo con instituciones que contaban con elevadas capacidades técnicas para formar sólidos procesos de macroevaluación educativa.

Pese al esfuerzo por institucionalizar la macroevaluación, en este periodo hubo una fuerte inestabilidad en relación con los organismos a cargo de la gobernanza de la macroevaluación, lo que finalmente obstaculizó el desarrollo de la innovación educativa que se pretendía. Entre 1986 y 1994 los procesos operaron bajo el convenio IIMEC-UCR-MEP con el fin de que se generara una confluencia entre la capacidad técnica y la logística para la implementación de las pruebas nacionales. Posteriormente, entre 1994 y 1996 se creó la Comisión Nacional de Bachillerato para la dirección y administración de la aplicación de los exámenes de bachillerato. Estuvo integrada por los siguientes actores del MEP:

Figura 3.8

Cambios institucionales en el desarrollo histórico de la macroevaluación educativa en Costa Rica



a/ Antes de 1944, conocido como la Secretaría de Instrucción.
Fuente: Barquero, 2023 con base en Francis, 2023a.

un delegado personal del Ministerio de Educación como presidente, la dirección de la División de Control de Calidad y Macroevaluación, la dirección de la División de Desarrollo Curricular, la dirección del Departamento de Asesoría para la Evaluación de Aprendizajes, un delegado de la Asesoría Técnico-Jurídica y un representante de la entidad especializada de los campos de medición y evaluación como representante del convenio IIMEC-UCR.

Entre 1996 y 1998, a través del Decreto Ejecutivo No. 20614-MEP, se creó el Centro Nacional de la Evaluación para el Aprendizaje (CENE-EDU) como un órgano desconcentrado del MEP adscrito al CSE bajo el cual se acogió el Programa de Pruebas Nacionales. Sus funciones eran organizar, desarrollar y ejecutar las acciones de evaluación del sistema educativo formal en sus diversas etapas y modalidades, con carácter nacional y para propósitos diagnósticos, formativos y sumativos, además de asumir la elaboración, administración y

calificación de las pruebas, así como la evaluación de diversos proyectos y programas educativos. El centro funcionó hasta 1998 y se cerró debido a la ilegalidad del decreto que lo creó.

Desde entonces y hasta la actualidad, la gobernanza de la macroevaluación recae estrictamente en el MEP. Específicamente, entre 1998 y 2007 estuvo a cargo de la División de Control de Calidad y Macroevaluación del Sistema Educativo. A partir del 2008, esta división se convierte en la Dirección de Gestión y Evaluación de la Calidad (DGEC), con dos instancias a cargo: el Departamento de Evaluación y Certificación (a cargo de las Pruebas Nacionales) y el Departamento de Evaluación de la Calidad (o Departamento de Supervisión Nacional) con una visión más asesora que de control e inspección.

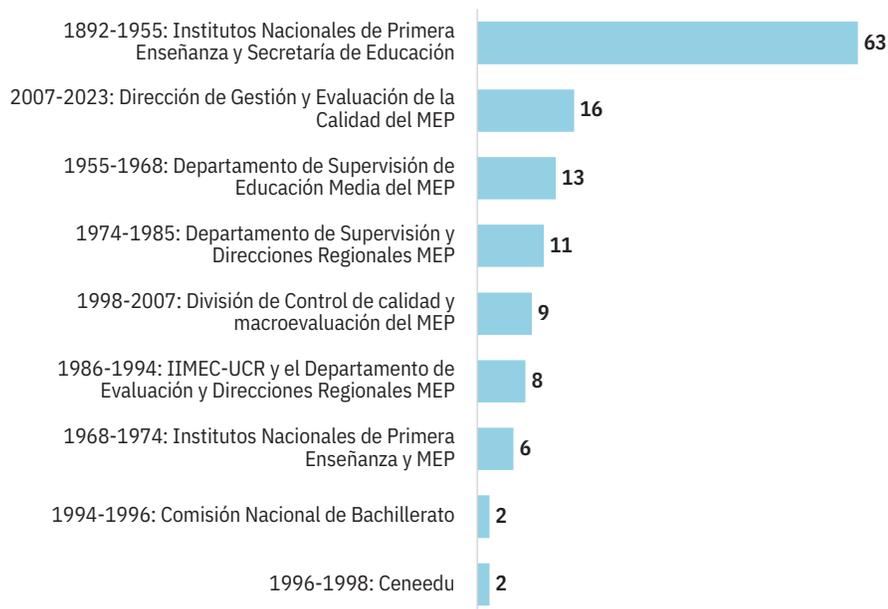
Queda claro que, a pesar de que el MEP ha asumido un protagonismo relevante, la instancia específica ha presentado una alta rotación a cargo, según la época del desarrollo de la macroevalua-

ción educativa en el país. El período de mayor estabilidad en la gobernanza se presentó entre 1892 y 1955 cuando estuvieron a cargo los Institutos Nacionales de Primera Enseñanza y la Secretaría de Educación (que posteriormente se convertiría en el MEP). Tras varias décadas de cambios repentinos, en el periodo de 2007 al 2023 la Dirección de Gestión y Evaluación de la Calidad asumió las funciones de macroevaluación en el sistema. Por otra parte, los años con menor estabilidad se dieron entre 1994 a 1998, cuando la Comisión Nacional de Bachillerato y el Ceneedu asumieron estas funciones dos años cada una (gráfico 3.1).

Ante este panorama, se resalta la importancia de definir una entidad competente que permita asegurar el rigor técnico, la independencia, la continuidad, la innovación y los recursos financieros necesarios para garantizar la viabilidad a largo plazo del sistema de macroevaluación y el alcance de sus objetivos. La gobernanza, por su parte, debe constituir un conjunto orgánico y articulado de

Gráfico 3.1

Instituciones a cargo de la macroevaluación educativa costarricense según los años de duración



Fuente: Barquero, 2023 con base en Francis, 2023a.

instituciones, procesos, instrumentos y acciones que permitan el análisis de los datos de la evaluación y contribuyan a las recomendaciones de políticas orientadas a promover una mejora educativa continua.

El conflicto de interés institucional amenaza la permanencia de la macroevaluación

El conflicto de interés institucional a lo interno del MEP consiste en que esta entidad ejecuta la política y a la vez evalúa sus resultados; es decir, ambas funciones se encuentran bajo la responsabilidad de un mismo jerarca. Esto contribuye a explicar también las modificaciones constantes en la organización institucional de la evaluación, así como los objetivos y funciones de la macroevaluación dentro del sistema. Debido a esta simbiosis, los procesos de evaluación se han caracterizado por su vulnerabilidad e inestabilidad ante ciertas coyunturas. A manera de ejemplo, es posible identificar tres ocasiones en las que se interrumpieron las acciones de macroevaluación en momentos críti-

cos cuando el sistema precisaba de información exacta y de alta calidad para mejorar los deficientes desempeños educativos que se estaban experimentando.

La primera y más extensa interrupción sucedió en la década entre 1974 a 1985, época en la que los estudios académicos reportaban una situación crítica en el rendimiento académico de los estudiantes. Este hecho generó una discusión nacional acerca de la necesidad de contar con mecanismos de transparencia y rendición de cuentas que permitieran mejorar el rendimiento de los estudiantes y velar así por el cumplimiento de los currículos vigentes. La segunda interrupción evaluativa ocurrió entre 2018 y 2020, como resultado de la pérdida de un semestre lectivo debido a una huelga y posteriormente a la afectación de la pandemia, momento en el cual se redujo cerca del 50% de los contenidos totales de los programas de estudio. La última interrupción se dio en 2022 debido a cuestionamientos técnicos sobre las pruebas FARO, sucesoras de las Pruebas Nacionales de Bachillerato, que inicialmente llevaron a su suspensión temporal

y finalmente a su cancelación sin que mediara una auditoría por parte del MEP o de un ente técnico externo. Para el 2023, el CSE retomó las evaluaciones estandarizadas, cuyo diseño y aplicación estaban en proceso a la hora de publicarse este *Informe*.

En todas las fases históricas analizadas, el vacío más importante se presenta en el aprovechamiento y uso de la información evaluativa para generar las transformaciones educativas necesarias. En este sentido, no ha bastado con que esto se encuentre normado y reglamentado para lograr la consolidación de mecanismos que integren los resultados, concreten estrategias de uso de dicha información en las aulas y gesten los cambios requeridos (Francis, 2023a). No se trata, entonces, de un resultado al azar, sino más bien de un efecto estructural en un problema básico que se resume de la siguiente forma: una evaluación subordinada directamente a los responsables políticos del MEP siempre quedará a merced de las necesidades y contingencias del gobierno de turno.

En conclusión, la estructura organizativa del MEP no ha logrado los flujos necesarios para aprovechar la información que resulta del proceso evaluativo, pues no existe claridad ni mecanismos para lograr que la macroevaluación funcione como un sistema de mejoramiento de la calidad. Además, aunque se observa que el sistema de macroevaluación se ha empleado como una herramienta para inspirar o legitimar reformas educativas, los objetivos planteados no estuvieron acompañados por la formulación inicial de instrumentos y alcances y con ello se contribuyó a desdibujar la importancia y eficacia de la evaluación en los procesos de mejoramiento de la calidad.



Para más información sobre

GOBERNANZA Y EL PROCESO DE MACROEVALUACIÓN EN COSTA RICA

Véanse Francis, 2023a, en www.estadonacion.or.cr

Gestión y evaluación de la calidad con la menor inversión educativa

La institucionalidad y la gobernanza de la macroevaluación educativa no solo han presentado problemas en su estabilidad y conformación, sino que, además, enfrentan limitaciones financieras para su adecuado funcionamiento y desarrollo. En el caso costarricense, la administración de los recursos financieros forma parte del presupuesto nacional destinado a la educación pública. La inversión destinada a los procesos de gestión y evaluación de la calidad es la menor de todos los programas presupuestarios del MEP. Entre 2019 y 2023, el presupuesto en este rubro ha representado en promedio un 0,17% del presupuesto total (MEP, 2023).

El contexto fiscal adverso que enfrenta el presupuesto de educación pública en los últimos años, como se evidencia en el capítulo 2 de este Informe, se ha traducido también en una disminución de recursos destinados a los procesos de gestión y evaluación de la calidad y con ello al sistema de macroevaluación nacional. A partir del 2020, se ha experimentado una reducción de la inversión en educación en este rubro y, en especial, en el 2023 se experimentó el decrecimiento más significativo de los últimos 10 años que alcanzó el -23% (gráfico 3.2).

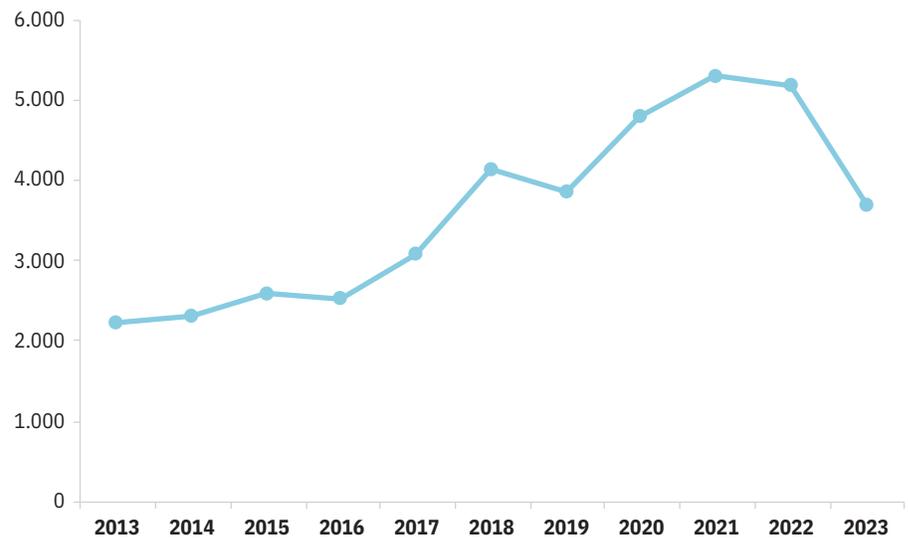
Deterioro progresivo del sistema de macromedición de aprendizajes en Costa Rica

La macromedición de los aprendizajes ha experimentado durante la última década un retroceso significativo que se ha caracterizado por cuestionamientos técnicos y rigideces al articularse con la política curricular. Esta situación ha generado debates sobre la pertinencia y relevancia de las mediciones estandarizadas para proveer insumos confiables que apunten a la mejora continua de los aprendizajes.

En particular, desde la reforma educativa llevada a cabo a partir del 2012, cuando los programas de estudios pasaron de un enfoque de contenidos a uno que enfatiza en el desarrollo de habilidades y competencias, el sistema educativo costarricense no ha logrado

Gráfico 3.2

Presupuesto destinado a los procesos de gestión y evaluación de la calidad en el Ministerio de Educación Pública. 2013-2023
(millones de colones)



Fuente: Barquero, 2023 con datos del MEP.

diseñar instrumentos de medición estandarizados acordes con los nuevos enfoques de enseñanza. Esta sección analiza el caso particular de Matemáticas y enfatiza en las debilidades de los ítems empleados para medir constructos complejos, como el desarrollo de la competencia matemática, asociados a capacidades cognitivas superiores como el planteamiento y la resolución de problemas.

Modificaciones metodológicas en las pruebas estandarizadas no han logrado generar mediciones robustas de los aprendizajes

Los especialistas consultados para la elaboración de este capítulo consideran que la existencia de pruebas nacionales democratiza el sistema educativo al establecer un estándar mínimo de aprendizaje en todas las instituciones del país. Aunque en Costa Rica las pruebas estandarizadas iniciaron de manera muy temprana en el sistema, han sufrido una serie de cambios en su permanencia, tipos, enfoques y pesos en la promoción de los estudiantes que, como se ha discutido, ha impedido la consolidación de un sistema de macroevaluación robusto.

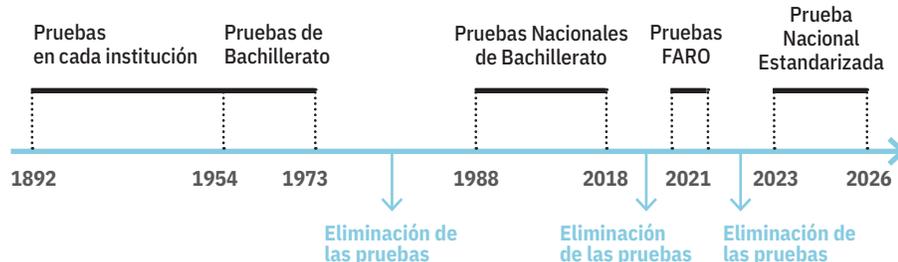
La figura 3.9 presenta la evolución de las pruebas que han caracterizado al sistema de medición de los aprendizajes en Costa Rica desde 1892 hasta la actualidad. El país ha realizado al menos cinco aplicaciones estandarizadas con distintas metodologías y propósitos de aplicación. A excepción de la Prueba Nacional Estandarizada (MEP, 2023), en cada caso se aplicaron pruebas para cada una de las asignaturas con el objetivo de medir las habilidades propuestas en los planes de estudio. Todas se han caracterizado por tener como finalidad la certificación de los aprendizajes alcanzados por la población estudiantil.

Entre 1892 y 1972 se implementaron dos pruebas para certificar la culminación de la Educación Media. La primera se implementó entre 1892–1954 y permitía la certificación del Bachillerato en Ciencias o Letras a cargo de los Institutos Nacionales de Segunda Enseñanza y Estudios Normalistas. La segunda se aplicó entre 1954 y 1973 y certificaba el Bachillerato en Ciencias o Letras a cargo del MEP.

Posteriormente, las Pruebas de Bachillerato se retomaron en 1988 y se

Figura 3.9

Trayectoria de las mediciones estandarizadas aplicadas en Costa Rica. 1892-2026



a/ Las pruebas de bachillerato aplicadas entre 1945 y 1973 eran exámenes de grado que certificaban el Bachillerato en Ciencias o Letras y de Maestro Normal. Por otra parte, las pruebas FARO corresponden a Pruebas para el Fortalecimiento y Renovación de los Aprendizajes.

Fuente: Barquero 2023 con base en Poveda et al., 2023 y Francis, 2023a.

Recuadro 3.4

Prueba Nacional Estandarizada 2023

En 2023 el CSE aprobó la propuesta del MEP para restaurar las evaluaciones estandarizadas en el país luego de la suspensión de las pruebas FARO ese mismo año. Para esto, durante el 2023 se aplicará la Prueba Nacional Estandarizada cuyo propósito es “monitorear el desarrollo de aprendizajes esenciales de los estudiantes para la mejora continua en el proceso de enseñanza y aprendizaje, entre el inicio y final del año académico, asimismo, para la medición de la calidad educativa (CSE, 2023)”.

Esta nueva evaluación se caracteriza por lo siguiente:

- Consiste en una única prueba que evalúa las asignaturas de Español, Matemáticas, Ciencias, Estudios Sociales y Educación Cívica, según los enfoques y fundamentos teóricos de los programas de estudio vigentes. Cabe destacar que Educación Cívica solo se evaluará en primaria.
- Se divide en dos aplicaciones: una aplicación diagnóstica que se llevó a cabo en abril, y una aplicación sumativa y de

certificación que se realizará durante el segundo semestre del año. La segunda aplicación es un requisito para obtener el certificado de conclusión de la Educación General Básica en el I y II Ciclo, así como el Título de Bachiller en Educación Media.

- Se aplica en formato digital y físico. El formato físico se utilizará en los centros educativos que presenten limitaciones a acceso a internet o recursos electrónicos.

Es importante señalar que, al momento de realizar el Informe, el PEN no había recibido respuesta al oficio CNR-PEN-033-123 del 22 de febrero del 2023 en el que se hacía una solicitud formal de información y aclaraciones técnicas sobre la elaboración y desarrollo de la prueba. No obstante, la respuesta fue recibida el 25 de mayo en el oficio CSE-SG0240-2023, por tanto, al momento de elaborar este Informe no fue posible realizar un análisis riguroso y detallado sobre la información suministrada por la entidad.

Fuente: CSE, 2023.

mantuvieron vigentes hasta 2018. De acuerdo con Francis (2023a), su objetivo era controlar la calidad y acreditación de los logros básicos, así como restablecer, readecuar y valorar la calidad de la educación. Estas pruebas se aplicaron con carácter sumativo y con fines de certificación para el último año de la secundaria¹². Durante este periodo, también se aplicaron pruebas diagnósticas para los niveles de tercero, sexto y noveno año con el objetivo de implementar medidas correctivas en los niveles iniciales.

En 2019, el MEP tomó la decisión de eliminar las pruebas FARO, pruebas que fueron diseñadas con el objetivo de avanzar de un modelo de medición de contenidos a uno de evaluación de habilidades e incorporar cuestionarios de factores de contexto para comprender las diferencias en el rendimiento (MEP, 2019). A diferencia de sus antecesoras, la propuesta inicial proponía una aplicación en dos momentos: la primera en el año previo al final de la etapa escolar o secundaria (décimo año para colegios académicos y undécimo año para colegios técnicos) y una segunda en el último año (sexto año, undécimo año y duodécimo año). Se pretendía que, con la información suministrada en la primera evaluación, se tomaran decisiones que permitieran mejores desempeños en la segunda evaluación. No obstante, ante las dudas técnicas realizadas a los cuestionarios de contexto, estas pruebas fueron suspendidas y solo se aplicaron en una primera etapa en 2021. Las evaluaciones eran de carácter sumativo y, por primera vez, sus resultados se entregaban con la descripción de los niveles de desempeño que alcanzaba cada estudiante, con el objetivo de informar sobre los aprendizajes esperados en los programas de estudio.

Finalmente, en 2023, la DGEC propone la implementación de la Prueba Nacional Estandarizada. A diferencia de las pruebas anteriores, esta consistirá en dos aplicaciones: al inicio y al final de sexto año, undécimo año y duodécimo año, según corresponda. Al momento de realizar el Informe, no se había recibido por parte del MEP información que permitiera efectuar un análisis técnico y riguroso de este instrumento (recuadro 3.4)

En este contexto, además de los frecuentes cambios metodológicos, deben recordarse las interrupciones en la macromedición de aprendizajes ocasionadas por las administraciones de turno que decidieron acerca de su suspensión o eliminación. Aunque cada interrupción tuvo una duración variable, juntas acumulan cerca dos décadas en las cuales el sistema educativo no contó con información sobre los desempeños de los estudiantes y por ende la política educativa careció de una base sólida para su diseño y ejecución.

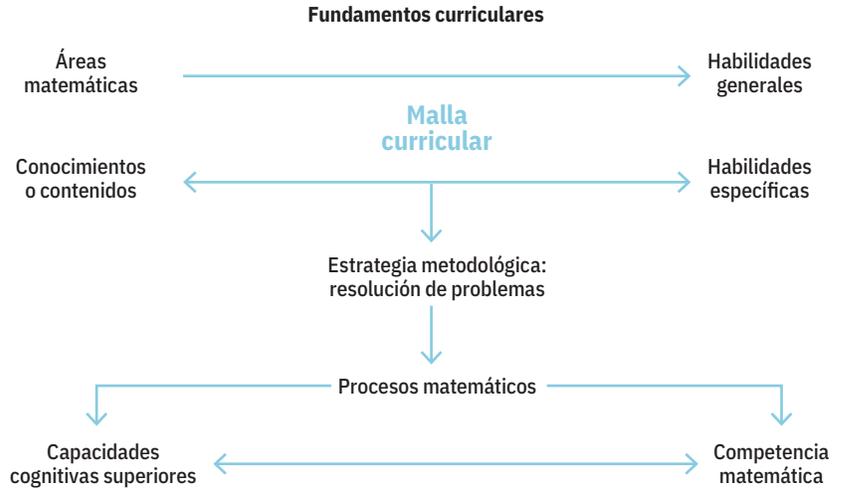
Desvinculación entre las pruebas de Matemáticas y los programas de estudio vigentes cuestiona la validez de la información obtenida sobre el desempeño de los estudiantes

Según Unesco (2015), una evaluación de alta calidad, pertinente y capaz de producir mejoras en el sistema educativo es aquella que opera en completa armonía con el currículo. Si no están sincronizadas, su eficacia se ve amenazada; se corre el riesgo de que el currículo sea secuestrado por normas o estrategias de evaluación engañosas o contradictorias y que la evaluación sea irrelevante o brinde señales imprecisas perjudiciales. El sistema educativo debe, por tanto, evaluar todo su currículo, incluyendo los contenidos, la forma en que se enseña, la capacidad de aplicar el contenido para resolver problemas y desempeñarse eficazmente en la vida social y personal e incluir las competencias personales que la sociedad espera que el sistema educativo ayude a los niños y jóvenes a cultivar.

Las investigaciones muestran que, desde el cambio curricular gestado a partir del 2012, en el que los programas de estudio pasaron de un enfoque de contenidos hacia el desarrollo de habilidades para el logro de competencias, el país no ha logrado realizar pruebas en correspondencia con los programas de estudios en áreas clave como Matemáticas. Un trabajo realizado por Poveda et al. (2023) para esta edición del Informe, analizó en detalle el caso de la macroevaluación de los aprendizajes en Matemáticas, una de las materias más relevantes en cuanto

Figura 3.10

Elementos curriculares de los Programas de Estudio de Matemáticas 2012 del MEP^{a/}



a/ En el proceso intervienen cinco ejes disciplinares: resolución de problemas, contextualización activa, actitudes y creencias positivas hacia las matemáticas, uso de las tecnologías y uso de la historia como elemento didáctico.

Fuente: Barquero, 2023 con base en Ruiz, 2018 y Poveda et al., 2023.

al cambio curricular que experimentó, con un modelo que incluye capacidades cognitivas superiores y de competencia general, en la que se han observado consistentemente los desempeños educativos más bajos.

Desde esta perspectiva, la implementación de los programas de estudio aprobados desde 2012 supuso un enfoque disruptivo para la enseñanza tradicional, ya que obligó a transformar los procesos de planeamiento, mediación y evaluación. Sin embargo, tras más de una década, el sistema educativo aún no ha logrado diseñar pruebas estandarizadas articuladas con la política curricular, lo que ha impedido determinar con precisión las habilidades y destrezas que están desarrollando los estudiantes. Esto, a su vez, obstaculiza la generación de insumos fundamentales para fortalecer la práctica docente y la toma de decisiones prontas y pertinentes para mejorar la acción en el aula y alcanzar los objetivos establecidos en los programas de estudio actuales (MEP, 2012).

En el caso de Matemáticas, una de las principales dificultades ha sido superar la organización de la evaluación horizontal entre contenidos, objetivos e ítems de las

pruebas estandarizadas anteriores a esta reforma, vigentes hasta el 2018. Según Poveda et al. (2023) y Ruiz (2018), esta organización contribuye a una desarticulación entre los conocimientos matemáticos aprendidos y los demás objetos curriculares de los programas.

Esta desarticulación obstaculiza el desarrollo de la competencia matemática, entendida como la herramienta que permite la comprensión y acción de la persona en la resolución de problemas en diferentes contextos de la realidad ciudadana (incluyendo científicos y matemáticos). Dicha competencia se construye con base en los conocimientos y sus habilidades asociadas, así como en capacidades cognitivas superiores, tales como razonar y argumentar, plantear y resolver problemas (figura 3.10). De acuerdo con Ruiz (2018), se requiere la implementación de una estrategia que funja como articuladora del proceso. Al respecto, una alternativa por considerar es la Estrategia 4+6, cuyo detalle se presenta en el recuadro 3.5.

Para el diseño de la evaluación, resulta necesario contemplar los conocimientos disciplinares teóricos, así como las habilidades específicas y generales que

Recuadro 3.5

Estrategia 4+ 6: modelo para la valoración de tareas matemáticas de acuerdo con el programa de estudios

La Estrategia 4+6 es una propuesta diseñada por Ruiz (2018) para la valoración de tareas matemáticas que pueden emplearse en la construcción de aprendizajes, la evaluación de aula y la macroevaluación. Como parte de la valoración y el diseño de tareas matemáticas, integra los diversos objetos curriculares: conocimientos, contextos, habilidades generales y específicas, procesos y sus grados de intervención y los niveles de complejidad de los problemas establecidos por los programas (reproducción, conexión, reflexión). Los grados de los procesos y los niveles de complejidad se establecen con base en modelos teóricos precisos que incluyen más de 60 indicadores y criterios de valoración (Ruiz, 2018). La Estrategia articula mediante una metodología precisa todos los elementos curriculares y la fundamentación de los programas. Comienza con el planteamiento del enunciado y la solución o soluciones de un problema; asimismo, incluye cuatro pasos y seis elementos (figura 3.11).

Esta propuesta se ha puesto en práctica en los diferentes recursos virtuales que ha elaborado en los últimos años el Proyecto de Reforma de la Educación Matemática en

Costa Rica en todos los ítems y prácticas de evaluación de las diferentes áreas matemáticas en Bachillerato, Bachillerato por Madurez, FARO, Recursos Libres de Matemáticas, Minimoocs, entre otros. También se ha incorporado en el diseño de problemas para ejemplificar la puesta en práctica del currículo. Estos materiales se han dirigido tanto a estudiantes como a docentes de matemáticas de primaria y secundaria, con mayor énfasis en esta última (<https://www.reformamatematica.net/>). Poveda et al. (2023) consignan resultados de la aplicación de esta estrategia en sus investigaciones.

Algunas de las ventajas identificadas de la Estrategia 4+6 son:

- Permite realizar un análisis completo que favorece la integración de habilidades (un aspecto destacable de estos programas).
- Permite identificar el contexto del problema, lo que contribuye su valoración.
- Permite valorar eficientemente el nivel de participación de los procesos matemáticos en la solución o soluciones del problema mediante el uso de los indicadores.

- Ofrece una herramienta efectiva para clasificar un problema de acuerdo con su nivel de complejidad, lo que permite también dosificar la presencia de los diferentes niveles según los requerimientos del diseño o la evaluación.
- Permite identificar si el problema en cuestión puede ser utilizado en la mediación pedagógica para una primera etapa (generar conocimiento nuevo) o para una segunda etapa (movilización y aplicación de los aprendizajes).
- Permite fundamentar el uso de diferentes problemas o tareas matemáticas que involucren una variedad elementos curriculares.

En suma, esta estrategia constituye una propuesta coherente e integradora que permite visualizar la participación de los fundamentos teórico-curriculares en el diseño de tareas, la evaluación de aula y la macroevaluación.

Fuente: Ruiz, 2018 y Poveda et al., 2023

deben alcanzar los estudiantes, las cuales se asocian a las áreas matemáticas del Programa (números, geometría, medidas, relaciones y álgebra, y estadística y probabilidad). Además, se deben tomar en consideración cinco procesos matemáticos que funcionan como elementos articuladores, a saber: razonar y argumentar, plantear y resolver problemas, conectar, comunicar y representar. Estos procesos deben potenciarse sistemática y reiteradamente a través de la acción de aula, sin distinción de los conocimientos y habilidades que se estén implementando.

Una evaluación pertinente demanda medir de forma adecuada las capacidades cognitivas superiores a las que aspira el programa; sin embargo, su alcance

Figura 3.11

Pasos y elementos de la Estrategia 4+6



Fuente: Ruiz, 2018.

requiere de un adecuado diseño y selección de los problemas que se vayan a utilizar en la micro y macroevaluación de aprendizajes. El programa vigente señala que para ello se necesita trabajar problemas en tres niveles de complejidad: reproducción, conexión y reflexión (MEP, 2012), los cuales involucren la activación de cada uno de los procesos matemáticos. Cada problema debe considerar los conocimientos específicos del área matemática por evaluar, las habilidades específicas y generales involucradas, el nivel de participación de cada proceso matemático y el nivel de complejidad del problema, todo esto sin descuidar el papel de los ejes disciplinares propuestos. De acuerdo con Ruiz (2018), una prueba nacional estandarizada que esté articulada con el currículo de matemáticas debe considerar los siguientes elementos (figura 3.12).

El trabajo realizado por Poveda et al. (2023) también revela otras limitaciones. A partir del 2016, progresivamente, la Prueba Nacional de Bachillerato tuvo modificaciones con respecto a lo que se venía implementando: se incorporó el área de probabilidad y estadística, se incluyeron ítems de respuesta corta y se empleó un mismo contexto para responder varias preguntas, entre otros cambios. No obstante, se mantuvo la tabla de especificaciones propia del sistema de pruebas vigente desde el año 2006 y que resultaba parcialmente contradictoria con el currículo aprobado en 2012.

En este sentido, tanto las pruebas de bachillerato como las pruebas FARO presentaban carencias en el desarrollo de los ítems de Matemáticas para medir adecuadamente lo establecido en el programa. En el caso de la prueba de bachillerato, los autores encontraron que los

ítems eran preponderantemente de selección única y tenían poca coincidencia con las habilidades generales propuestas en el programa, además de que carecían de la incorporación de los procesos de resolución de problemas y de la contextualización activa.

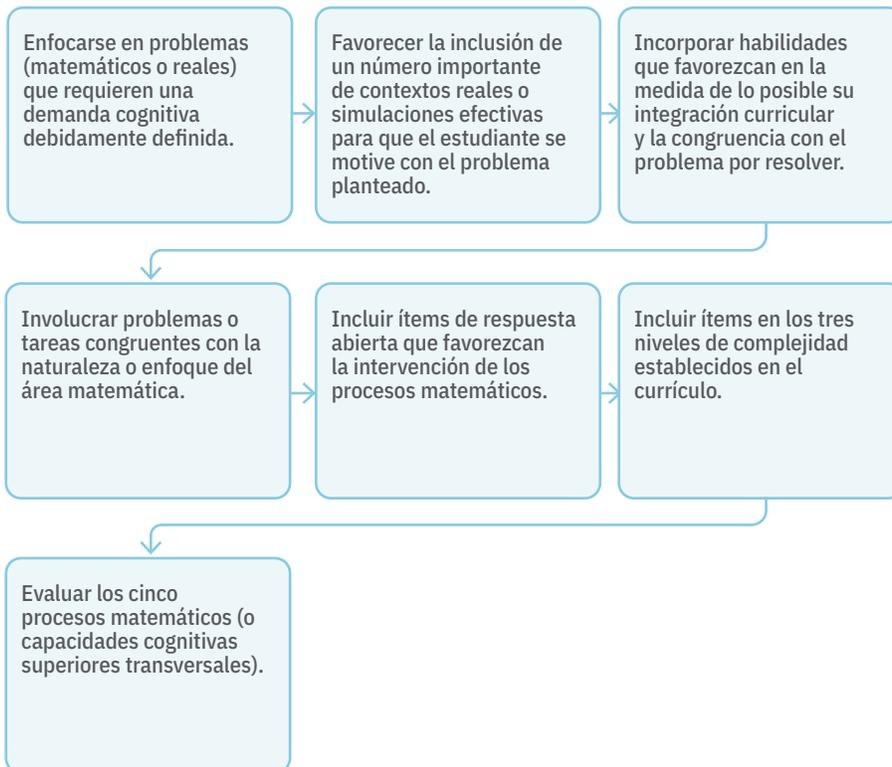
En relación con las pruebas FARO, se introdujeron algunos elementos novedosos con el objetivo de mejorar la concordancia entre los programas de estudio y las pruebas. Uno de estos aspectos fue la inclusión de un ítem de respuesta construida, además del mantenimiento de las preguntas de respuesta breve que ya se habían incorporado en años previos en el bachillerato. Asimismo, se incluyeron problemas que presentaban un contexto real, lo que permitía medir las habilidades específicas y generales del estudiante al aplicar los conocimientos básicos de las funciones para resolver la situación presentada.

Sin embargo, la prueba seguía incorporando predominantemente ítems de selección única y algunos de estos ítems partían de contextos artificiales que no aportaban información relevante para la solución del ítem. Tanto en las PNB como en FARO no se incluyó la integración de habilidades ni tampoco se planteó evaluar el desarrollo de los cinco procesos matemáticos, elementos cruciales de los programas de Matemáticas. Cabe recalcar que, finalmente, al momento de elaborarse este Informe, aun no se contaba con información en detalle sobre la Prueba Nacional Estandarizada de 2023, por lo que no se puede realizar un análisis de esta.

Si no se solventan los problemas señalados, los procesos macroevaluativos generarán una serie de consecuencias adversas. La primera de ellas implicaría la obstaculización de los procesos que permiten integrar, planear, medir y evaluar dentro del currículo matemático vigente (Poveda et al., 2023). La segunda es que, si el contenido, enfoques y las capacidades sobre las que se examina a los estudiantes no se ajustan al currículo, la mayoría de los docentes no se apegará a los programas en su mediación pedagógica, pues normalmente enseñará para la prueba (*teaching for the test*).

Figura 3.12

Elementos clave para la elaboración de una prueba estandarizada congruente con el programa de estudios de Matemáticas



Fuente: Barquero, 2023 con base en Ruiz, 2018 y Poveda et al., 2023

Estas consecuencias repercutirán negativamente sobre los estudiantes en varios sentidos. A nivel cognitivo, este tipo de enseñanza limitada priva a los estudiantes de la oportunidad de profundizar y consolidar su comprensión conceptual, de contenido o técnica, mediante la exclusión de oportunidades para explorar el tema desde una variedad de ángulos y utilizarlo de manera pertinente. Además, disminuye el entusiasmo del estudiante para estudiar y aprender, pues convierte la curiosidad y la ambición en rutina y memoria (Unesco, 2015). Finalmente, debido a los vacíos en la definición de enfoques adecuados sobre los conocimientos (y habilidades), así como en la conciencia acerca del papel relevante que tienen los contextos reales y el cultivo de los cinco procesos matemáticos fundamentales para el desarrollo de la competencia matemática, no se puede pretender que se desarrolle esta competencia general, máxime cuando en las aulas y en la macroevaluación no se consideran.

Para más información sobre

CRITERIOS Y RUTAS PARA EL DISEÑO DE PRUEBAS NACIONALES DE MATEMÁTICAS
véase Poveda et al., 2023
www.estadonacion.or.cr

Costa Rica lejos de implementar prácticas efectivas de macroevaluación educativa

Esta sección analiza prácticas internacionales efectivas propias de los sistemas educativos que han logrado consolidar sistemas exitosos de macroevaluación de aprendizajes. Se enfoca exclusivamente en la macroevaluación de los desempeños estudiantiles y se ilustra cómo se pueden gestionar de manera efectiva los factores centrales del proceso, a saber: su organización institucional y formas de gobernanza, sus diferentes objetivos, las técnicas e instrumentos empleados para medir los aprendizajes estudiantiles

y el uso que se hace de los resultados obtenidos para mejorar la calidad de los aprendizajes.

La conclusión principal de este análisis es que el país se encuentra considerablemente lejos de implementar las prácticas más efectivas que se llevan a cabo a nivel internacional en esta área. Esto impide el uso eficiente de los recursos y la disponibilidad de información de calidad en momentos críticos en los que es urgente mejorar la calidad educativa.

Autonomía técnica caracteriza la organización institucional de los sistemas de macroevaluación

La literatura especializada coincide en que la práctica más pertinente en materia de organización institucional es la creación de una entidad o agencia especializada dedicada al diseño y aplicación del proceso macroevaluativo. La recomendación es que se trate de una entidad autónoma respecto de las autoridades educativas para evitar situaciones donde se es “juez y parte” en la evaluación. Ello implica que dicha entidad debería tener independencia financiera y política e, idealmente, contar con una fuerte presencia de especialistas entre sus directivos.

Esta autonomía en macroevaluación empezó a desarrollarse en el mundo a partir de la década de los noventa, pero se consolidó en los últimos años. Desde entonces, se han identificado varias ventajas. En primer lugar, se fortalece la rendición de cuentas pública de las autoridades en cuanto a los resultados del sistema educativo. En segundo lugar, se logra una “masa crítica” de experticia en la materia, lo que resulta fundamental para desarrollar una evaluación cada vez más adecuada y precisa. En tercer lugar, permite la consolidación de la capacidad institucional para aplicar políticas de macroevaluación. Por último, al independizar la evaluación educativa de los ministerios de educación se evitan o aminoran los vaivenes abruptos en este campo por razones políticas o ideológicas derivadas de cambios en la administración gubernamental, tal como ocurrió en Costa Rica con la suspensión de las pruebas FARO en 2022 (Rodino, 2023a).

En América Latina existen ejemplos de autonomía en la macroevaluación. Existe en países como Chile, Uruguay, Colombia y Perú (cuadro 3.4). En todos estos casos, se resaltan las ventajas de contar con una entidad autónoma o al menos con importantes niveles de autonomía para impulsar de manera constante la evaluación educativa. La consecuencia positiva es un avance cada vez mayor en desarrollos técnicos y capacidades institucionales, así como en resultados robustos a nivel evaluativo e impactos positivos en los aprendizajes estudiantiles (Rodino, 2023a).

Además de los casos latinoamericanos resaltados, Rodino (2023a) lista 21 países más donde existen agencias nacionales de evaluación educativa, 16 de ellas en países europeos y el resto en América, Asia y Oceanía¹⁴.

Definición de propósitos en los sistemas de macroevaluación con equilibrios entre fines sumativos y formativos

En cuanto a los marcos teóricos y normativos de la evaluación, la recomendación internacional tiene que ver con la búsqueda de equilibrios entre los objetivos formativos y sumativos (reseñados en la figura 3.2), para lo cual se requiere contar con un sistema que cumpla todas las funciones de la evaluación, dado que un único modelo o evaluación no permite alcanzar todos los objetivos. Por ello, resulta conveniente aplicar simultáneamente un tipo de evaluación que identifique las mejoras necesarias en la enseñanza —usualmente a lo interno de las escuelas— y otro tipo para calificar y rendir cuenta de los aprendizajes alcanzados, que corresponde, por lo general, a macroevaluaciones externas.

En los sistemas educativos considerados de mayor calidad (en Europa y Asia) siempre hay al menos un instrumento de evaluación asociado a la promoción o avance cuando el estudiantado concluye la secundaria (Ruiz, 2018). La mayoría de los países miembros de la OCDE confían las evaluaciones de certificación en un sistema mixto que combina evaluaciones continuadas realizadas por los docentes durante la secundaria y un examen

Cuadro 3.4

Características de la institucionalidad y en la gobernanza a cargo de la macroevaluación en cuatro países latinoamericanos

Características	Países			
	Chile	Uruguay	Colombia	Perú
Autonomía institucional	Establecida por ley.	Establecida por ley.	Adscrita a la autoridad educativa.	Dependencia de la autoridad educativa.
Autonomía financiera	Cuenta con presupuesto propio.	Cuenta con presupuesto propio.	Cuenta con presupuesto propio.	No tiene presupuesto propio
Cargos / funcionariado	Concursables.	Directiva nombrada por el gobierno, funcionarios son privados.		Nombramientos internos.
Evaluación de logros educativos mediante pruebas estandarizadas.	✓	✓	✓	✓
Amplia difusión a los actores del sistema para el desarrollo de políticas de mejoramiento.	✓	✓	✓	✓
Revisión permanente de componentes de evaluación.	✓		✓	✓
Desarrollo de investigaciones sobre resultados.	✓	✓	✓	
Brindar orientación a centros educativos.	✓	✓	✓	
Producción de conocimientos para toma de decisiones.	✓	✓	✓	
Desarrollo de estándares educativos, indicadores de desempeño, así como criterios y modalidades de valuación.	✓	✓	✓	
Evaluación del cumplimiento de metas y objetivos del sistema educativo.	✓	✓		
Asesoría sobre evaluaciones internacionales.	✓	✓		
Ayudar a construir una narrativa positiva hacia la evaluación.			✓	✓
Asistir a otros agentes del sistema en sus procesos internos de evaluación.		✓		

a/ La existencia de la función se indica con la marca con “✓”.
Fuente: Vargas, 2023 con base en Rodino, 2023a.

al final del ciclo que se diseña y procesa externamente (Rodino, 2023a). El recuadro 3.6 sintetiza las principales funciones de las pruebas de certificación externa.

Existe una amplia variedad en los tipos de pruebas de certificación que los distintos países llevan a cabo en relación con sus propósitos y con los momentos en los que se aplican (cuadro 3.5). Cabe resaltar que en Europa prevalece la prueba de certificación como requisito de graduación de secundaria, la cual se promedia con las calificaciones obtenidas en la escuela, en tanto que en los países asiáticos no se realiza ninguna prueba de conclusión, sino una de admisión muy estricta

para quienes optan por realizar estudios superiores.

Recientemente, en América Latina se han generado arduos debates sobre las pruebas de certificación debido a la existencia de brechas en la calidad de la educación recibida por estudiantes de sectores sociales más vulnerables, zonas rurales y modalidades educativas más complejas (como la educación nocturna). Este hecho recalca la necesidad de un acompañamiento sistemático de este tipo de evaluaciones con acciones pedagógicas complementarias y compensatorias que tomen en cuenta la disponibilidad de recursos, infraestructura y

personal docente, áreas en las que, a menudo, flaquean los centros educativos que muestran desempeños más bajos. En la actualidad, las pruebas de certificación solo se aplican en República Dominicana y Colombia.

Medición estandarizada alineada al currículo con inclusión de factores de contexto y aprovechamiento de las tecnologías

Existe una gran cantidad de procedimientos, técnicas e instrumentos que se pueden emplear en las macroevaluaciones para satisfacer los objetivos que

Cuadro 3.5

Pruebas externas de certificación que se realizan en diversos países, según objetivos y momento en que se aplican

Momento de aplicación	Objetivo	Prueba y países
Al finalizar la secundaria	Medir el desempeño estudiantil.	Prueba de Matriculación en Finlandia. Prueba de Bachillerato en Francia.
	Retroalimentar al sistema educativo.	Prueba de Madurez en: Austria, Bosnia y Herzegovina, Bulgaria, Croacia, Hungría, Italia, Kosovo, Liechtenstein, Macedonia, Montenegro, Polonia, República Checa y Ucrania.
Al optar por el ingreso a la universidad.	Fungir como instrumento para admitir a la población estudiantil a la educación terciaria.	Prueba en Japón, Corea del Sur y China.

Fuente: Vargas, 2023 con base en Rodino, 2023a.

Recuadro 3.6

Principales funciones de las macroevaluaciones para certificación

Las pruebas de certificación brindan una clausura o cierre a una etapa de la preparación de estudiantes que han llegado a su madurez, cerrando un proceso formativo y abriendo nuevos desafíos. En este sentido, las funciones de estas pruebas son:

- Garantizar la labor que cumplió el sistema educativo preuniversitario, aportando transparencia acerca del valor de los certificados educativos ante la sociedad.
- Apoyar la cultura evaluativa de un país, lo que resulta importante y productivo para el sistema educativo.
- Aportar evidencias e información clave al personal docente y su entorno escolar para la correcta implementación del currículo oficial.
- Aportar orientaciones sobre prácticas educativas.
- Servir de tamiz para la educación superior.

Fuente: Rodino, 2023a.

se persiguen. El escenario ideal consiste en alcanzar en la mayor medida posible los resultados deseados en tanto se evitan o minimizan las limitaciones e inconvenientes. En su investigación sobre buenas prácticas internacionales, Rodino (2023a) señala cuatro aspectos fundamentales para lograr una adecuada integración entre los objetivos de la evaluación y las técnicas e instrumentos utilizados:

- Alinear la evaluación con los objetivos curriculares y los estándares educativos.
- Investigar los factores asociados a los resultados de la evaluación (como se hace en las pruebas ERCE (Estudio Regional Comparativo y Explicativo, de Unesco) y PISA, en las que ha participado Costa Rica.
- Incorporar Tecnologías de Información y Comunicación (TIC).
- Considerar experiencias y aportes internacionales relacionados con las macroevaluaciones.

Macroevaluación alineada a los objetivos curriculares y estándares educativos

La evaluación estudiantil debe estar, en principio, claramente alineada con los objetivos del currículo y sus estándares educativos¹⁵, ya que estos determinan el tipo de instrumentos a utilizar y sus características. No es posible diseñar una

evaluación basada en un currículo centrado en contenidos de la misma manera que en otro que se basa en el desarrollo de habilidades o competencias.

En muchos países, existe un énfasis creciente por desarrollar y emplear estándares educativos ambiciosos como bases de la evaluación y la rendición de cuentas. Su adopción posibilita medir el desempeño estudiantil contra un resultado deseable, lo que permite controlar la calidad con más facilidad y logra que los resultados de la evaluación sean más comprables. Este movimiento hacia la comparación de los resultados estudiantiles con estándares educativos también ha motivado a diversos países a adoptar macroevaluaciones estandarizadas (Rodino, 2023a). Ejemplos de países que implementaron estándares educativos nacionales son Australia, Austria, Bélgica, Dinamarca, Francia, Luxemburgo, Noruega, Reino Unido y Estados Unidos. Cabe aclarar que los países que adoptaron este sistema lo hicieron de manera flexible, adaptándolos a su contexto y cultura. En consecuencia, los estándares educativos varían en especificidad y énfasis entre países (Rodino, 2023a).

Investigar factores del contexto asociados con los resultados

Los factores asociados al desempeño estudiantil son estudios que se realizan simultáneamente a la aplicación de pruebas de macroevaluación. Usualmente, se basan en cuestionarios complementarios que recogen información sobre las características socioeconómicas y culturales

del estudiantado y sus familias, el funcionamiento de los centros escolares a los que asisten y la metodología pedagógica con la que se imparten las asignaturas específicas. Con base en esta información, se puede identificar cuáles datos inciden significativamente, ya sea de forma positiva o negativa, en los resultados de las pruebas. Esta información es muy relevante para las autoridades educativas, pues permite orientar el diseño de políticas públicas que fortalezcan los factores favorables y debiliten o desincentiven aquellos que son desfavorables para el logro de aprendizajes estudiantiles.

Las pruebas estandarizadas internacionales¹⁶ de Unesco y OCDE en las que el país ha participado han indagado sobre los contextos de aprendizaje. En el caso de ERCE 2019, se recopiló información sobre cuatro actores educativos: estudiantes (sus características, actitudes y percepciones); familias (características económicas, demográficas y culturales, del barrio y la escuela, así como las expectativas educativas sobre sus hijos e hijas); docentes (características, antecedentes, prácticas, expectativas sobre los y las estudiantes, liderazgo y gestión escolar); y directores de escuela (características personales, infraestructura y gestión escolar). El estudio de Rodino (2023a) presenta ejemplos del tipo de preguntas que se hacen en los cuestionarios de factores asociados.

Por su parte, PISA¹⁷ se orienta de manera significativa hacia el impulso de políticas educativas que permitan mejorar los sistemas escolares a través del aporte de datos coyunturales y longitudinales que permitan identificar brechas y desarrollar las rectificaciones necesarias. Por ello, conecta los resultados de las pruebas de los estudiantes con datos sobre sus características familiares, socioeconómicas, culturales, prácticas de estudio y otros factores clave que inciden en su aprendizaje, tales como la formación docente, su metodología, el nivel de compromiso del profesorado, sus creencias y su estabilidad laboral, además de las principales características del centro escolar. Estos factores asociados se recaban mediante cuestionarios que acompañan a la materia priorizada en

la prueba de ese año de aplicación. Los cuestionarios son respondidos por estudiantes, docentes y autoridades escolares (Rodino, 2023a).

El Estado de la Educación ha realizado varias investigaciones sobre los resultados de las pruebas PISA (PEN, 2013; 2015, y 2017) y en 2016 conformó un equipo con el MEP para analizar los principales factores identificados para Costa Rica en la prueba PISA de 2015. Las pruebas de factores asociados no se aplican únicamente en macroevaluaciones internacionales. En Latinoamérica, Chile, Colombia y Uruguay incluyen este tipo de análisis en sus pruebas nacionales, al igual que lo hacen otros miembros de la OCDE como Canadá, la Comunidad Flamenca de Bélgica, Suecia y Australia (Rodino, 2023a).

Por las consideraciones anteriores, los fallos técnicos en el diseño de la sección de factores asociados en las pruebas FARO de 2021 pueden acarrear severas consecuencias para la política educativa. En efecto, de persistir las resistencias políticas y sociales para su inclusión en las mediciones, el país se privará de un conocimiento indispensable para el diseño de estrategias de aprendizaje adaptadas a las características específicas de entornos socioterritoriales y de las poblaciones estudiantiles y docentes.

El beneficio de incorporar las TIC en las macroevaluaciones: la experiencia internacional y el caso de la prueba de dominio lingüístico de la UCR

El crecimiento de la macroevaluación con pruebas estandarizadas y el manejo de los numerosos datos que genera se han visto muy beneficiados por la gran capacidad de las TIC, lo que ha contribuido a su generalización y empleo cada vez más eficiente. El adecuado uso de las TIC introduce muchas mejoras en el proceso, como por ejemplo el hecho de que aumenta la posibilidad de realizar evaluaciones más individualizadas, de forma más rápida, a mayor escala, con más confiabilidad y a un costo menor. Asimismo, permite valorar de una forma más acertada las habilidades cognitivas como la resolución de problemas.

La sistematización de experiencias realizada por Rodino (2023a) identificó cinco ejemplos exitosos en cuanto al uso de medios digitales en la aplicación de pruebas de diversa índole (cuadro 3.6). Tres de estos casos se ubican en América Latina (Costa Rica, con la Prueba de Dominio Lingüístico del inglés de la UCR, Uruguay y Colombia) y dos en Europa (Noruega y Dinamarca). Estos ejemplos son evidencia del valor de las TIC para difundir los resultados entre agentes que pueden tomar decisiones clave a nivel de aula y de sistema educativo y para el acopio de información relevante que permita desarrollar investigaciones relacionadas con algunos de los siguientes temas: factores que inciden en los resultados, análisis de cohortes, impacto de métodos pedagógicos, condiciones escolares o de gestión de los centros educativos sobre el aprendizaje estudiantil, entre otros.

En esta línea destaca, a nivel nacional, el acopio de un banco de ítems y pruebas desarrolladas localmente por la Escuela de Lenguas Modernas de la Universidad de Costa Rica (ELM-UCR) producto de la investigación constante de expertos en el campo. En los casos expuestos resalta, asimismo, la rapidez con la que se pueden obtener resultados, la facilidad para agregar resultados e identificar patrones, brechas y necesidades (en los cinco países), la posibilidad de realizar adecuaciones a las condiciones del contexto o de los propios estudiantes y de dar seguimiento a su progreso en el tiempo (como en el de Uruguay, Colombia y Dinamarca) y la flexibilidad para integrar no solo pruebas estandarizadas sumativas, sino también pruebas formativas a nivel de aula (como en el ejemplo noruego).

Cabe destacar que, en el caso de la ELM-UCR, se realizan investigaciones y publicaciones sobre los resultados y se promueve la profesionalización continua del equipo de trabajo al incentivar la continuación de sus estudios de posgrado. Actualmente, se encuentran desarrollando, además, una prueba para evaluar la producción oral en inglés a través del uso de inteligencia artificial¹⁸ (Rodino, 2023a).

Por otra parte, en Colombia se está

Cuadro 3.6

Buenas prácticas en la incorporación de las TIC en los procesos de macroevaluación

País - Evaluación	Entidad ejecutora	Población evaluada	Forma en que incorpora y usa las TIC
Costa Rica Prueba de Dominio Lingüístico del Idioma Inglés, con el objetivo de fortalecer la enseñanza de este idioma. Se aplica desde 2009/2010."	Escuela de Lenguas Modernas-Universidad de Costa Rica	Docentes	Desarrollo de plataformas para alojar exámenes y herramientas de apoyo.
		Estudiantes universitarios	Posibilidad de ver en tiempo real quién está realizando la prueba.
		Estudiantes de primaria y secundaria	Aplicación de pruebas en diferentes dispositivos electrónicos.
		Posibilidad de ampliar a toda la probación gracias al uso de las TIC.	Obtención rápida y simultánea de los resultados nacionales y regionales.
			Ofrece pruebas con adecuaciones curriculares.
Uruguay Pruebas Aristas (estandarizadas muestrales).	Instituto Nacional de Evaluación Educativa	Estudiantes de ciertos grados de primaria y secundaria.	Aplicada integralmente por computadora.
			Complementada por esfuerzos de autoridades educativas para cerrar brechas de conectividad y acceso a recursos digitales (Plan Ceibal).
			Resultados generados rápidamente y compartidos con tomadores de decisiones, medios de comunicación y la opinión pública.
			Una plataforma digital que permite acumular datos a lo largo del tiempo. Se puede acceder a ellos en cualquier momento (Arista en clase).
			Los docentes pueden generar una prueba en clase y contrastarla con resultados de estudiantes de otros centros en condiciones similares.
			Los resultados llevan recomendaciones para las personas docentes.
Colombia – Pruebas SABER.	Instituto Colombiano para la Evaluación de la Educación	Desde educación inicial hasta educación superior	Existe la exigencia de que los resultados de las pruebas se suban rápidamente a la plataforma informática.
			Se hace hincapié en que los resultados estén disponibles pronto para los tomadores de decisiones educativas.
			El sistema tiene capacidad para personalizar los aprendizajes a fin de adecuarlas a las necesidades de los alumnos.
Dinamarca Pruebas nacionales, desde 2010.		Estudiantes de primaria y secundaria	Pruebas se aplican íntegramente por computadora.
			Hay flexibilidad para adaptar los ítems a los niveles de habilidades de los estudiantes y tomando en consideración su desempeño previo.
			Retroalimentación a docentes al día siguiente de realizada la prueba.
Noruega Pruebas varias, desde 2008.		Estudiantes de primaria y secundaria	Todas las primarias y secundarias del país utilizan el sistema de pruebas administradas por medios electrónicos.
			El sistema se puede utilizar para aplicar pruebas nacionales estandarizadas, así como evaluaciones formativas más breves y rápidas.
			Maneja módulos de valor agregado para determinar influencia del centro educativo y la persona docente sobre el aprendizaje estudiantil.
			El sistema de datos fomenta el intercambio de información entre los agentes del sistema educativo.

Fuente: Vargas, 2023 con base en Rodino, 2023a.

empezando a promover que la herramienta elaborada por funcionarios del INEED se emplee en la formación docente, algo que es posible dado que la última ley de gobernanza de la institución eliminó la privacidad de los datos sobre estudiantes, docentes y escuelas para investigaciones educativas (Ospina y Ángulo, 2023).

Aprovechar la información que brindan las pruebas internacionales para mejorar la calidad de los aprendizajes

Los debates educativos al interior de los países sobre los resultados de los y las estudiantes están cada vez más influenciados por comparaciones entre naciones, las cuales ocurren gracias a la difusión de los resultados de pruebas internacionales como PISA, PIRLS, TIMSS y ERCE (Rodino, 2023a). Algunos sistemas educativos han experimentado fuertes sacudidas al conocerse los puntajes comparativos internacionales, lo que, en algunos casos, ha motivado la introducción de reformas educativas. Por ejemplo, a raíz de los resultados de PISA, se evidenció que el sistema educativo danés, uno de los que cuenta con la mayor inversión educativa en el mundo, obtenía resultados promedio muy similares a los de otros países de la OCDE que cuentan con menores niveles de inversión. Por su parte, Francia realizó una reforma para introducir el núcleo común de competencias educativas.

Tanto Dinamarca, como Italia, México y Portugal desarrollaron pruebas nacionales estandarizadas que están en consonancia con los objetivos nacionales de aprendizaje. Algunas de esas pruebas se basaron metodológicamente en las evaluaciones internacionales y, en particular, en el marco propuesto por PISA. Esto ha implicado un aumento en el intercambio de experiencias entre países para medir los resultados de aprendizaje.

Al respecto, Rodino (2023a) advierte que un país no necesariamente debe copiar procedimientos, técnicas o instrumentos de macroevaluación utilizados en las pruebas internacionales o en las de algún otro país sin un análisis de su pertinencia y adecuación en el contexto

local. Sin embargo, cultivar la disposición y el interés libre de prejuicios por conocer otras experiencias que ocurren fuera de las propias fronteras ayuda a valorar aprendizajes que podrían ser útiles.

Difusión y uso de resultados para la promoción de políticas educativas basadas en evidencia

El último de los factores clave en el proceso de macroevaluación se refiere a las estrategias para asegurar que sus resultados se conozcan masivamente y tengan efectos sobre las acciones que se realizan en aulas y escuelas del sistema educativo con el fin de mejorar los desempeños estudiantiles. El uso de los resultados de las evaluaciones estandarizadas corresponde a una de las principales lagunas identificadas por personas expertas latinoamericanas consultadas por Rodino (2023a), quienes concuerdan en que, pese a la robustez de los instrumentos aplicados y a la amplia disposición de información, se emplean de forma escasa o deficiente los sus resultados.

En este aspecto, también se identificaron buenas prácticas internacionales en cuatro aspectos específicos: la comunicación de resultados, la disponibilidad de la información, su uso en aulas y la promoción de políticas públicas. Los detalles de esas buenas prácticas se desarrollan a continuación en los siguientes acápite.

Amplia difusión de los resultados hacia la opinión pública y distintos agentes de la comunidad educativa

Informar o comunicar los resultados es un propósito clave de la evaluación que adquiere cada vez más importancia. Resulta fundamental para sensibilizar a la opinión pública y a las personas tomadoras de decisión dentro y fuera del sistema educativo acerca de la situación del estado de la educación en el país.

Por otra parte, la amplia difusión de los resultados de las evaluaciones nacionales estandarizadas denota un compromiso con la transparencia y la rendición de cuentas, lo que aumenta la confianza en el sistema. Sin embargo, quizá lo más relevante es que brinda información y

orientación que permite sustentar las acciones tendientes a mejorar la calidad de la educación.

Los esfuerzos de divulgación deben orientarse hacia varios niveles:

- Al sistema educativo en general y a la opinión pública, mediante reportes sobre el estado de la educación y la compilación de bases de datos educativas a partir de las cuales se pueden desarrollar análisis relevantes.
- A los centros educativos, a través del acceso a sitios web con información o reportes anuales de inspecciones, entre otras herramientas.
- Al estudiantado y sus familias, mediante los reportes de las calificaciones y del desempeño por escuelas.

Irlanda del Norte, Chile y Australia representan ejemplos eficaces en cuanto al uso que le dan distintos actores a la información de las macroevaluaciones educativas. En el primer caso, el país aumentó las inspecciones a instituciones de educación secundaria y los reportes de esas inspecciones se publican (algo que sucede igualmente en Inglaterra). De la misma forma, se han introducido pruebas estandarizadas obligatorias en primaria. Adicionalmente, se ha promulgado una estrategia nacional para la alfabetización lingüística que introduce estándares nacionales de logros estudiantiles y obliga a brindar a padres y madres información evaluativa adecuada y clara para que comprendan el progreso de sus hijos e hijas.

Por su parte, en Chile los resultados de las pruebas SIMCE enriquecen la información del sistema con *vouchers* escolares. En Australia se publican los resultados del desempeño individual de los centros educativos en un sitio web de fácil acceso para el público en general que permite comparar los resultados entre instituciones y territorios. Este tipo de informes públicos de desempeño de los centros educativos es muy importante también en Bélgica (Comunidad Flamenca), Dinamarca, Inglaterra, Islandia y en la República Eslovaca (Rodino, 2023a).

Poner los resultados a disposición de las personas docentes y capacitarles para su análisis e interpretación

La difusión de los resultados a los diversos actores del sistema educativo no garantiza que la información se emplee para detonar los cambios requeridos en la mejora del desempeño estudiantil. El que dichos resultados se traduzcan en la renovación de las prácticas escolares (a nivel de estrategias, recursos, técnicas o materiales) también está en función de que las personas docentes y quienes desempeñan las funciones administrativas de los centros sepan interpretar y aprovechar la información macroevaluativa. Por ello, se requiere apoyar el desarrollo de capacidades en docentes y administradores para fomentar la comprensión y el uso eficaz de esos resultados y para estimular su compromiso de participar activamente en discusiones de evaluación externa.

En este sentido, destacan dos ejemplos internacionales: Canadá e Italia. En la provincia canadiense de Ontario se realizó una campaña de alfabetización evaluativa a nivel de escuelas y distritos. Esta capacitación se enfocó en desarrollar capacidades profesionales de docentes y administrativos para el análisis de datos, el establecimiento de expectativas sobre el uso de los estándares educativos, el apoyo a las escuelas para entender y analizar sus datos específicos y la capacitación a las personas docentes para traducir resultados evaluativos en estrategias y planes de mejoramiento con metas claras. También se enfatizó en la importancia de retroalimentar a los agentes del sistema acerca de estos análisis para contribuir a un mejoramiento global.

Por su parte, en Italia, el Instituto Nacional para la Evaluación del Sistema Educativo (Invalsi) desarrolla acciones a fin de promover entre docentes y directores de escuelas el uso de los resultados de las macroevaluaciones estandarizadas. Invalsi crea un “documento de identidad de la escuela” que sintetiza el desempeño de cada centro educativo en la macroevaluación y aporta datos contextuales sobre los resultados estatales, regionales y provinciales como un medio para motivar

diálogos entre docentes, directores e inspectores sobre lo que se puede aprender de los resultados estudiantiles. (Invalsi, 2010).

Uso regular de los resultados de la evaluación para producir mejoras en el aula

Uno de los puntos críticos de un sistema de macroevaluación es garantizar que sus resultados se vinculen directa y efectivamente con las prácticas de aula. Una forma de hacerlo es involucrando a las escuelas en prácticas de autoevaluación significativa y mejorando la capacidad del personal docente para evaluar formativamente a los estudiantes (Rodino, 2023a).

Si se desarrollan estructuras, procedimientos y recursos a nivel central con este fin, resulta necesario establecer roles claros para la implementación de las políticas de evaluación a nivel local. Se debe trabajar en conjunto con personas directoras y supervisoras de escuelas, así como con otras autoridades educativas locales. Un ejemplo en este sentido proviene del estado de *South New Wales* en Australia, que desarrolló un sistema de información mediante una herramienta de análisis de datos llamada Caja de Herramientas para la Medición, Evaluación e Información sobre las Escuelas (*SMART*, por sus siglas en inglés).

A través de este recurso, el personal docente puede enseñar a sus estudiantes, retroalimentar a los agentes escolares, publicar material de apoyo para otros docentes y mejorar la administración del conocimiento. *SMART* funciona como una red que conecta a maestros, maestras y escuelas con las mismas preocupaciones, y permite crear una mejor infraestructura de datos para la investigación educativa. El sistema *SMART* puede vincularse con los sistemas escolares de administración de datos que rastrean el progreso de estudiantes con el fin de seguir su desarrollo en el tiempo y compartir esta información con otros docentes o con la siguiente escuela a la que concurra el estudiante.

Por su parte, Noruega cuenta con un Portal de Escuela que permite monitorear la información y el análisis de sus propios

resultados. Este portal en internet ofrece resultados de aprendizaje, ambientes de aprendizaje, recursos y datos básicos sobre escuelas. El portal cuenta con una sección accesible al público en general en la que se brinda información sobre los resultados educativos nacionales. Otra área del portal se encuentra protegida por una contraseña y en ella las escuelas pueden consultar información más detallada que les permite compararse con el promedio nacional y así fundamentar sus procesos de autoevaluación.

La Comunidad Francesa de Bélgica cuenta con un portal similar (el Portal de la Enseñanza) que además brinda al público información extensa sobre objetivos de aprendizaje del estudiantado, indicadores y legislación educativa, así como las herramientas pedagógicas e instrumentos necesarios para que las escuelas se involucren en evaluaciones externas.

Promoción de políticas educativas basadas en evidencias de la macroevaluación

En el ámbito internacional, se documentaron tres tipos de medidas tomadas por los países para recolectar evidencia que respalde la toma de decisiones y el desarrollo de políticas educativas a nivel nacional y local: la creación de instancias o instituciones especializadas, la formación de alianzas con otras entidades para consolidar la evidencia y establecer sistemas de monitoreo y la realización de investigaciones y análisis para enriquecer la evidencia y su publicación (figura 3.13).

La opción de contar con una entidad específica o bien de encargar la tarea de analizar las evidencias generadas por las macroevaluaciones institucionales existentes depende de las preferencias y capacidades de los diversos países. Un tema en el que se encontró amplia coincidencia es en la promoción a nivel internacional de investigaciones de fondo para informar sobre la toma de decisiones y el diseño de políticas educativas. Por ejemplo, la Comunidad Francesa de Bélgica propone al gobierno un plan de investigación multianual en el que se establecen

Figura 3.13

Medidas adoptadas en otros países para promover políticas educativas basadas en evidencias de la macroevaluación

Creación de instituciones para investigación, asesoramiento y difusión

- Consejo Asesor de la Academia Neozelandeza de Valoración, con participación de académicos e investigadores.
- Consejo para la Evaluación y el Desarrollo de Calidad de la Educación Primaria y Secundaria Inferior.
- Comisión de Monitoreo (Comunidad Francesa de Bélgica).
- Instituto Federal para la Investigación Educativa, la Innovación y el Desarrollo del Sistema Escolar (BIEFE, por sus siglas en alemán) de Austria.

Alianzas estratégicas con entidades existentes

- Vinculación con universidades (Luxemburgo).
- Vinculación con centros o institutos de investigación independientes (Corea del Sur y Francia).

Desarrollo de investigaciones y análisis sobre:

- Mejores prácticas para apoyar el marco metodológico de evaluación (Nueva Zelanda y Dinamarca).
- Implementación (pilotaje) y uso de lo último en metodologías de evaluación (Dinamarca).
- Indicadores y estándares educativos (Nueva Zelanda, Austria).
- Análisis y monitoreo de los resultados de las evaluaciones (todos los países analizados).
- Identificación de buenas prácticas de aplicación de resultados en mejoramiento en aulas y escuelas (Nueva Zelanda).
- Identificación de brechas educativas y detección de necesidades (Comunidad francesa de Bélgica).
- El análisis del impacto de las políticas educativas y su implementación (Corea del Sur).
- El monitoreo del sistema educativo (Francia).

Fuente: Vargas, 2023 con base en Rodino, 2023a.

Para más información sobre

BUENAS PRÁCTICAS NACIONALES E INTERNACIONALES DE MACROEVALUACIÓN EDUCATIVA

véase Rodino, 2023a

www.estadonacion.or.cr

las prioridades de la investigación educacional y los objetivos que deben alcanzar esas investigaciones.

En varios países, el asunto de la difusión orientada a quienes toman decisiones sobre la política educativa es de gran relevancia y se encuentra directamente asociado con las labores de las entidades creadas y/o de las investigaciones desarrolladas. Por ejemplo, en Nueva Zelanda se

publica la Síntesis de Mejores Evidencias Reiteradas (BES, por sus siglas en inglés) del Ministerio de Educación, en el cual se analizan las investigaciones sobre factores escolares que tienen efecto positivo en el aprendizaje estudiantil. Estas publicaciones las utilizan diversos tomadores de decisión en cuanto a políticas y grupos de interés para alimentar la política y la práctica educativa en el país.

En Dinamarca, el Consejo para la Evaluación y el Desarrollo de Calidad de la Educación Primaria y Secundaria Inferior tiene entre sus mandatos el fortalecimiento del intercambio sistemático de conocimientos entre las instituciones de investigación y las escuelas para promover el uso del conocimiento por parte de estas últimas en aras de mejorar su enseñanza. Otro mandato de dicho Consejo es el análisis y divulgación de resultados investigativos para apoyar la

creación de políticas orientadas a mejorar las escuelas. En Austria, el BIEFE tiene la responsabilidad de consolidar el sistema de monitoreo educativo y preparar informes nacionales de educación¹⁹.

Retos clave y portafolio de sugerencias para consolidar un sistema efectivo de macroevaluación educativa en Costa Rica

La evidencia presentada en este capítulo es concluyente: el país enfrenta importantes desafíos en todos los componentes que integran el sistema de macroevaluación educativa. Estas limitantes obstaculizan la puesta en marcha de transformaciones que aseguren el aprendizaje efectivo de los estudiantes. Resulta necesario tomar medidas concretas para superar estas dificultades. Sin embargo, esta tarea es compleja debido a la relativa

escasez de estudios especializados en nuestro país, así como a la dificultad de encontrar soluciones efectivas a causa del carácter sistémico de la macroevaluación y de la alta calidad técnica que debería contemplar cualquier propuesta que procure la mejora del actual estado de cosas.

Esta sección del capítulo adopta un enfoque propositivo para abordar cuatro problemas no resueltos de la política pública en macroevaluación, identificados por las investigaciones que fundamentan este capítulo. Se trata de un portafolio de sugerencias para enfrentar esos problemas, un menú de ideas, con la advertencia expresa de que no se trata de una “receta única” ni una hoja de ruta de política pública claramente definida. La selección e implementación de estas sugerencias requieren de un diálogo nacional y de la mediación entre los distintos actores educativos involucrados en la materia para adaptarlas a las condiciones actuales del sistema educativo costarricense.

Cada problema por resolver se relaciona con uno de los cuatro componentes del sistema de macroevaluación. Vistos en su conjunto, no son, evidentemente, los únicos que habría que abordar, pero todos constituyen asuntos importantes que, de atenderse, desencadenarían un importante progreso en relación con la situación actual.

Para cada uno de los cuatro problemas identificados se plantean vías de solución, consideraciones y recomendaciones basadas en las experiencias, aprendizajes y advertencias planteadas por los expertos consultados en Chile (Medeiros, 2023) Colombia (Ospina y Angulo, 2023), Perú (Miranda, 2023), Uruguay (Lasida, 2023; Miranda y Monroy 2023), así como en la extensa revisión de otras experiencias internacionales realizada por Rodino (2023b). Asimismo, se tomaron en cuenta las recomendaciones aportadas por expertos y personalidades educativas nacionales consultados. No obstante, cabe resaltar que el análisis, interpretación y organización de estos insumos es producto de la reflexión del equipo técnico del Programa Estado de la Nación, en quien recae la responsabilidad de las recomendaciones enunciadas.

El marco teórico y normativo para la evaluación educativa

Definición del problema de política pública a resolver

En Costa Rica no existe un marco teórico y normativo claro para el desarrollo de las macroevaluaciones educativas. Si no se tiene claro el marco de referencia, difícilmente se logrará sacar provecho incluso de las evaluaciones técnicamente mejor realizadas.

Consideraciones propositivas

Se requiere un acuerdo entre los distintos actores del sistema educativo y otros decisores políticos sobre:

- ¿Qué se quiere evaluar?
- ¿Para qué se quiere la evaluación? ¿Cuál es el o los propósitos de la evaluación y qué dimensiones se van a evaluar? Se pueden combinar propósitos, pero estos deben estar claramente establecidos (recuadro 3.7).
- ¿A quién se va a evaluar?
- ¿Con qué frecuencia se va a evaluar?
- ¿Cuáles aspectos se van a evaluar en forma continua y periódica y cuáles podrán modificarse según necesidades específicas o situaciones coyunturales?
- ¿Quién va a utilizar la información (públicos meta y niveles de devolución) y cuáles requerimientos deberían tener los resultados para ser pertinentes para esos actores? Es decir, ¿cuál va a ser la estrategia de difusión?
- Si se van a emplear estándares e indicadores educativos como base, se recomienda medir el avance o retroceso en las sucesivas evaluaciones. En el recuadro 3.8 se incluyen recomendaciones de los especialistas consultados con respecto a las características de los estándares educativos.

Delimitar este marco permitirá, a su vez, perfilar la gobernanza institucional del sistema que implementará las evaluaciones, así como precisar el tipo de evaluaciones por realizar, definir las características técnicas (y tecnológicas) de las evaluaciones para responder a los propósitos establecidos y establecer los métodos de recolección, procesamiento,

análisis, reporte y difusión de los resultados. Igualmente, incidirá sobre el tipo de monitoreo y seguimiento que se realice, tanto de los resultados obtenidos, como de la implementación de los propios procesos evaluativos.

La definición del marco de referencia debería estar acompañada por el establecimiento de una línea base de las capacidades nacionales para la evaluación educativa, según se detalla en la figura 3.14.

La literatura enfatiza en el requerimiento de que este marco cuente con un enfoque integral u “holístico” de la evaluación que permita responder a necesidades en distintos niveles del sistema educativo. En todos los casos, se menciona la importancia de contar con una variedad de enfoques evaluativos sumativos, formativos, así como de autoevaluaciones y coevaluaciones (Ospina y Angulo, 2023; Miranda y Monroy, 2023).

En su precisión, será importante establecer qué se entiende por ciertos conceptos frecuentemente utilizados como “integralidad” o “equilibrio”, para lo cual resulta necesario precisar no solo los enfoques evaluativos, sino también las poblaciones objetivo, las áreas de conocimiento que se pretende evaluar, si se van a cubrir los contenidos definidos en los currículos educativos, así como las competencias transversales o variables de contexto para los estudios de factores explicativos, entre otros aspectos. También se debe establecer qué se entenderá por “pertinencia” de los resultados, la cual es una característica deseable externada por todos los especialistas consultados que implica, en particular y entre otros factores, precisar los requerimientos necesarios para transmitir la información. Miranda y Monroy (2023) plantean que dicho marco debería pensarse en función de producir “resultados capitalizables” para la mejora del sistema educativo, tomando en cuenta las características y condiciones contextuales particulares de los procesos de enseñanza y aprendizaje.

En este sentido, una consideración relevante para su desarrollo implica posicionar al sistema de evaluación como un catalizador de la reflexión educativa que

Recuadro 3.7

Propósitos de la evaluación

Las personas expertas latinoamericanas consultadas mencionan diversos propósitos para la evaluación educativa, los cuales determinan el tipo de pruebas que se realizan, la metodología empleada, el tipo de análisis y de informes o reportes que se presentan a los distintos actores, así como los canales de comunicación utilizados para difundir los resultados, entre otros aspectos.

En general, todos los sistemas de evaluación apuntan hacia el mejoramiento de la calidad educativa, pero es en la precisión de este gran objetivo donde se observan diversos énfasis o combinaciones de propósitos (Rodino, 2023b).

Uno de los propósitos u objetivos principales se relaciona con la rendición de cuentas. En este punto se incluye desde dar cuenta de

la protección al derecho a una educación de calidad —según se plantea en los Objetivos de Desarrollo Sostenible vinculados con este ámbito— hasta monitorear la efectividad de las medidas, políticas e intervenciones adoptadas por el sistema educativo. Para lograr este propósito se requiere de información agregada.

Los receptores principales de la información serían las autoridades educativas, los tomadores de decisiones a nivel político, el público en general y los organismos internacionales que llevan el pulso de estas cuestiones.

Otro grupo de propósitos se relacionan con el fomento de procesos para el aseguramiento de la calidad de la educación de manera más específica. Algunos ejemplos son los siguientes: obtener información para acompañar a los centros educativos en la implementación de mejoras de estrategias pedagógicas que incidan en un mayor aprendizaje de sus

alumnos; propiciar investigaciones con este mismo fin; retroalimentar las estrategias de intervención en diversas regiones, tipos de centros o modalidades educativas; o aportar información clave para la formación inicial y continua de las personas docentes.

Miranda y Monroy (2023) advierten que la evaluación no debería concebirse como un mecanismo de control o penalización a centros educativos y docentes, especialmente si no cuentan con las herramientas, recursos y capacidades necesarias para realizar las mejoras que indique el análisis de los resultados de las evaluaciones.

Fuente: Vargas, 2023 con base en Rodino, 2023b y Miranda y Monroy, 2023.

Figura 3.14

Elementos para establecer una línea base de capacidades nacionales para la evaluación educativa

Diagnóstico de fortalezas y debilidades para la evaluación educativa (técnico-metodológicas y para la difusión de resultados).

Mapear fuentes nacionales e internacionales de conocimiento técnico sobre evaluación.

Estado del arte a nivel nacional sobre investigación de factores asociados con el aprendizaje educativo.

Mapeo de expertos en distintas áreas vinculadas con la evaluación (psicometría, estadística, manejo de datos, tecnología aplicada a procesos evaluativos, etc.).

Mapa de funciones, metas, plazos e interrelaciones vinculadas a las diferentes etapas de la evaluación (para acotar las funciones específicas del ente evaluador), así como normativa vigente sobre manejo de información sensible.

Identificar requerimientos de recursos, presupuesto, equipamiento, capacidades técnicas, tecnología y soporte para los procesos de difusión de resultados

Recuadro 3.8

Sobre los estándares educativos

La definición y empleo de estándares educativos como base tanto en el diseño como en el análisis de las pruebas macroevaluativas es una práctica extendida a nivel internacional (Rodino, 2023a). Los estándares corresponden a descripciones de lo que los estudiantes deben saber (estándares de conocimiento) y ser capaces de hacer (estándares de desempeño) en diferentes etapas del proceso de aprendizaje. Su uso implica un proceso en sí mismo que requiere de acuerdos, diseño, pruebas y revisiones. Medeiros (2023) indicó que en Chile la primera definición de estándares educativos tardó aproximadamente 2 años en concretarse, por lo que se desarrolló en paralelo al proceso de consolidación del sistema evaluativo de ese país.

Los estándares educativos deberían estar definidos en función del currículo de cada nivel y área por evaluar. La recomendación generalizada es que los resultados de las evaluaciones que se basan en ellos y se expresen en función de logros o déficits según los aprendizajes esperados en cada año escolar.

Estos estándares pueden ser útiles para la definición de indicadores, los cuales a su vez deben orientar a los evaluados sobre las mejoras o ajustes que deben realizar. Asimismo, los estándares permiten la definición de nuevas metas para el sistema educativo en

su conjunto. En este sentido, existe una amplia y valiosa experiencia internacional sobre el desarrollo y empleo de estándares e indicadores educativos que vale la pena considerar, como lo detalla Rodino (2023a).

Además, resulta esencial que los estándares consideren los perfiles de egreso y de progresión entre niveles educativos y al final del proceso de formación (Lasida, 2023). Sobre este último punto, se debe definir si el fin de las evaluaciones es la certificación o bien el aporte de insumos para el acceso a las etapas siguientes del sistema educativo, tal y como se ha empleado para el acceso a la educación terciaria, según mencionan Rodino (2023a), Miranda (2023) y Ospina y Angulo (2023).

En síntesis, los estándares e indicadores se pueden utilizar para:

- Identificar logros y brechas en los aprendizajes, es decir, lo que las personas evaluadas deberían saber en distintas fases de su proceso de enseñanza-aprendizaje.
- Medir logros y brechas con respecto a las competencias o habilidades cognitivas, metacognitivas o socioemocionales; dicho de otro modo, lo que las personas deberían estar en capacidad de hacer en el nivel en el que se les evaluó.
- Para informar los procesos de formación docente inicial o continua.

- Para formular diagnósticos y mapas de intervención prioritaria (Miranda, 2023; Medeiros, 2023), rendir cuentas o pasar a formar parte del currículo mismo (Lasida, 2023).

Además, resulta conveniente contar con estándares para distintos tipos de evaluaciones y poblaciones evaluadas que se hayan definido en el marco de evaluación (estudiantes, centros educativos, docentes, etc.).

Medeiros (2023) menciona algunos indicadores que se pueden desarrollar, tal como los Indicadores de Desarrollo Personal y Social, que incluyen aspectos relacionados con el clima de convivencia en los centros educativos, la participación y formación ciudadana, los hábitos de las personas evaluadas, la autoestima y la motivación, entre otros. Esta información se puede cruzar con datos administrativos del sistema educativo sobre matrícula, asistencia, retención, titulaciones y equidad de género, lo que aportaría profundidad a los análisis de resultados de la evaluación.

Fuente: Vargas, 2023 con base en Rodino, 2023a; Medeiros, 2023; Ospina y Angulo, 2023; Lasida, 2023, Miranda, 2023 y Miranda y Monroy, 2023.

promueva el diálogo permanente sobre los procesos pedagógicos. En suma, el objetivo se orienta hacia desarrollo de un sistema que aprende de las experiencias nacionales e internacionales y que devuelve resultados y análisis “útiles” para la toma de decisiones enfocadas en la mejora continua y en los distintos niveles (desde las prácticas pedagógicas en el aula hasta la definición de políticas educativas).

La definición de objetivos y propósitos es fundamental, de la misma forma que lo es la programación (orden y plazos) a partir de la que se concretarán las diversas evaluaciones. No todas las evaluaciones tienen que suceder al mismo tiempo, sino que se puede aplicar un criterio de progresividad. Por ejemplo, Lasida (2023)

advierte que, antes de desarrollar evaluaciones de docentes que puedan tener consecuencias sobre su permanencia laboral, resulta necesario ponderar las consecuencias de las brechas de los aprendizajes observadas en las macroevaluaciones de los estudiantes tomando en cuenta diversos factores concomitantes.



Institucionalidad y gobernanza

Definición del problema de política pública a resolver

En Costa Rica la instancia evaluadora está subordinada a uno de los principales actores del sistema evaluado (la cartera ministerial) y cuenta con poca autonomía e independencia técnica, económica y política para desarrollar su labor. Esta situación crea un conflicto de intereses y no ha generado incentivos para el desarrollo de las capacidades técnicas y operativas de la instancia evaluadora.

Consideraciones propositivas

Las personas expertas consultadas coincidieron en que para implementar el marco de referencia de la evaluación y que esta tenga un impacto sobre la calidad educativa en el país, es necesario contar con un acuerdo institucional que habilite la concertación, las definiciones fundamentales, el diseño e implementación de las recomendaciones que se deriven de ellas.

En el país, sin embargo, cuando se discute sobre evaluación, el énfasis recae casi exclusivamente en las pruebas que se realizan, lo cual impide articular el entramado de actores y de los arreglos institucionales necesarios para que funcione como un sistema nacional de evaluación. Estos arreglos incluyen al ente directamente encargado de implementar las evaluaciones, pero también lo trascienden (Ospina y Angulo, 2023). En este sentido, los esfuerzos deben dirigirse a desarrollar un modelo de gobernanza “distribuida” (Miranda y Monroy, 2023) que permita el accionar colegiado de los diferentes actores que intervienen en el proceso evaluativo (recuadro 3.9) en sus distintos ámbitos de acción, especialmente a nivel técnico.

Para lograr el objetivo, el sistema de macroevaluación requiere de un marco legal y procedimental concreto. Las funciones de los distintos actores que intervienen en el diseño, implementación, procesamiento, análisis y difusión de resultados deben aclararse en un instrumento legal estable, ya sea una ley o bien a nivel constitucional. De la misma forma, deben existir los reglamentos correspondientes para cada normativa.

Recuadro 3.9

Actores que considerar al definir la gobernanza del sistema

Existe una gran cantidad de actores que pueden intervenir en distintos procesos de un sistema nacional de evaluación educativa. Algunos son públicos, otros privados; unos son parte del sistema educativo, en tanto que otros son externos, aunque se relacionan con este. Algunos operan a escala nacional y otros a escala local.

Los principales actores identificados por los diversos autores consultados se dividen en dos grupos: actores que forman parte del sistema educativo y los externos relacionados.

- Actores que forman parte del sistema educativo: El ente evaluador, la población estudiantil, los padres y madres de familia, el profesorado, el cuerpo directivo y otro personal de los centros educativos, supervisores, asesores, autoridades educativas, otros departamentos del MEP (por ejemplo, la Dirección de Desarrollo Curricular y el CSE).

- Entes relacionados: Asamblea Legislativa, Centros de Investigación, universidades, expertos nacionales e internacionales en distintas ramas vinculadas con la evaluación, otros decisores de política pública (poder ejecutivo, Ministerio de Hacienda, entre otros), público en general.

Cada uno de estos actores cuenta con funciones propias, así como con formas de interacción e intereses particulares, los cuales en ocasiones se contraponen con los de otros actores. Por ejemplo, Miranda (2023) advierte que el sistema educativo tiene una estructura anidada, por lo que lograr la movilización de procesos dependerá de las capacidades, concepciones, contextos y cadenas de decisión, entre otros aspectos.

Fuente: Vargas, 2023 con base en Rodino, 2023b; Medeiros, 2023; Ospina y Angulo, 2023; Lasida, 2023, Miranda, 2023 y Miranda y Monroy, 2023.

Al precisar estas funciones, se debe tener en cuenta que el sistema de macroevaluación educativa es parte de uno más amplio: el sistema educativo. Debido a ello, es importante que considere los insumos y necesidades de los distintos actores de este último y que exista claridad de que habrá una diversidad de posiciones políticas en torno los asuntos evaluativos que requerirá de la implementación de procesos de diálogo y consulta. Adicionalmente, es recomendable que las funciones del ente a cargo del diseño e implementación de las evaluaciones se deslinden con claridad de las que puedan asumir otros actores con funciones relacionadas, como por ejemplo las instancias con orientaciones de política educativa, asesoría, desarrollo curricular, supervisión o conducción del proceso de enseñanza.

Según el consenso de los expertos consultados, se considera necesario establecer un ente evaluador que goce de autonomía en sus funciones, preferiblemente independiente del organigrama del MEP. La autonomía implica que dicho ente evaluador posea autoridad y legitimidad frente a los sujetos evaluados. Aunque se escuchen y consideren las necesidades y sugerencias de los distintos actores, se recomienda que los sujetos de evaluación posean una injerencia mínima en los aspectos técnicos de la evaluación y en el quehacer de la entidad en general.

Con base en este criterio, y tras consultar a personas expertas en legislación costarricense, se identificó que el país dispone de dos posibilidades: la creación de un órgano desconcentrado adscrito al MEP y la creación de una instancia independiente del MEP. Las características de

cada una de ellas se detallan en el cuadro 3.7. Independientemente de la figura que adopte este ente evaluador, debe cumplir con las siguientes características: transparencia, autonomía, flexibilidad, agilidad, comunicación, solvencia técnica, eficiencia, planificación y transparencia.

Que dicha instancia esté fuera del Ministerio de Educación Pública no significa que se desvincule totalmente del sistema educativo. Por el contrario, debe mantenerse en “buena sintonía” con las autoridades educativas, las cuales pueden participar en la junta directiva, comisiones asesoras, revisiones de planes, presentaciones de resultados, ejercicios de reflexión, procesos de mejora institucional, apoyo en temas curriculares o la definición de estrategias sobre cómo transmitir los resultados de las evaluaciones a los centros educativos y las aulas (Miranda, 2023).

Para lograr dicha autonomía, una de las sugerencias es que el ente evaluador disponga de independencia administrativa, operativa, financiera, técnica y de gestión. Además, el que cuente con un presupuesto propio y con la capacidad para establecer convenios de cooperación a nivel nacional e internacional en campos técnicos es vital, así como que cuente con la posibilidad de captar recursos externos para el desarrollo de proyectos o el intercambio de datos estadísticos con los centros o instituciones que los producen.

Asimismo, las personas expertas consultadas consideran que la selección de los máximos cargos directivos debe realizarse bajo concurso público con base en criterios de experiencia profesional y capacidades para desempeñarse en los puestos específicos. En adición, es importante que la entidad sea flexible al contratar o reorganizar equipos, con el fin de atender los objetivos de evaluación y reforzar áreas en donde se detecten puntos débiles. En caso de no contar con capacidades internas, puede considerarse la externalización de algunos servicios.

En aras de fortalecer su capacidad técnica, se recomienda que el ente evaluador mantenga relación con instancias especializadas en la materia a nivel internacional para que participen en la valoración de los procesos educativos

desarrollados en el país, así como para la actualización de su personal en psicometría, estadística, tecnología aplicada, inteligencia artificial, manejo de datos, entre otros. De igual modo, un ente evaluador eficaz establece alianzas con universidades y centros de investigación para apoyar en el diseño, evaluación de la implementación de las pruebas, procesamiento, análisis, generación de informes y reportes, capacitación de personal y de los usuarios de la información, así como en la formación inicial y continua de las personas docentes.

En relación con la eficiencia en la gestión, la entidad podría explorar opciones costo-eficientes que permitan responder a los propósitos evaluativos, aprovechar las evaluaciones internacionales en las que el país participa y complementarlas con otros tipos de evaluación que permitan profundizar y aportar datos más específicos para actores del sistema que operan a nivel micro o local. Sería conveniente buscar la complementariedad de instrumentos en lugar de yuxtaponerlos (Rodino, 2023b) y su aplicación mesurada para no convertirse en una carga que desplace actividades pedagógicas y genere “fatiga evaluativa”.

Además, se sugiere contar con una estructura “ligera”, no burocrática, lo suficientemente ágil para navegar los procesos complejos de toma de decisiones. Para superar la excesiva burocracia, es recomendable contar con instancias pequeñas de dirección y procesos institucionales sencillos. Además, las áreas de trabajo podrían organizarse de distintas formas (figura 3.15). En lo que concuerdan autores y autoras es en la importancia de contar con un consejo consultivo técnico estable, conformado por personas académicas y expertas nacionales e internacionales y, de ser necesario, comisiones más específicas para apoyar en determinadas áreas. Miranda y Monroy (2023) señalan que, además de considerar las capacidades técnicas, estas instancias brindan una buena oportunidad para procurar representación de diversas visiones políticas. Otro punto de coincidencia se refiere a la necesidad de contar con equipos dedicados a la comunicación social de los resultados de las evaluaciones.

Resulta aconsejable desarrollar ejercicios de planificación estratégica a corto, mediano y largo plazo, con base en metas e indicadores definidos en función del modelo evaluativo acordado y considerando las distintas fases del proceso de evaluación educativa. Debe existir rendición de cuentas sobre procesos y resultados, así como apoyo a otros actores del sistema educativo en la rendición de cuentas.

Figura 3.15

Opciones de organización del ente evaluador

Áreas divididas en función de las etapas del proceso evaluativo

- Diseño de instrumentos, aplicación y validación.
- Implementación y difusión.
- Monitoreo y seguimiento.
- Administración y gestión.

Organización tradicional

- Dirección Ejecutiva.
- Comunicación social.
- Evaluación (a cargo del proceso completo).
- Tecnologías de Información y Comunicación.
- Mejora educativa.
- Planeación.
- Administración

Fuente: Vargas, 2023 con base en Rodino, 2023b; Medeiros, 2023; Ospina y Angulo, 2023; Lasida, 2023; Miranda, 2023 y Miranda y Monroy, 2023.

Cuadro 3.7

Posibilidades para la creación de un ente evaluador autónomo, según la legislación costarricense

Características	Órgano desconcentrado adscrito al MEP	Instancia independiente del MEP
Regulación	<p>La desconcentración administrativa es una técnica de traslado de competencias dentro de una misma estructura administrativa. Se asigna una competencia específica a un órgano inferior debido a su especialidad y se le excluye, en cuando a esa competencia, de la línea de mando del jerarca.</p> <p>Es posible realizar la desconcentración mediante decreto ejecutivo, lo cual no requiere de un proceso de aprobación legislativa, aunque sí de un proceso para creación de actos generales, regulado en la Ley General de la Administración Pública.</p> <p>Requiere de un proceso de reestructuración orgánica del MEP, dentro del cual Mideplan emite su aprobación.</p>	<p>Al crearse una nueva persona jurídica, debe aprobarse por ley de la República.</p> <p>Asimismo, deberá justificarse técnicamente su creación y establecerse la fuente de financiamiento para su operación.</p> <p>Deberá decidirse qué tipo de ente será: una institución autónoma, semiautónoma o de otra naturaleza. La decisión incidirá en el procedimiento legislativo que se deba seguir.</p>
Independencia	<p>La desconcentración no brinda absoluta independencia funcional al órgano al depender orgánicamente del MEP, sino solo en cuanto a la competencia técnica que se le asignó en forma exclusiva, siendo el o la ministra el superior jerárquico de dicho órgano.</p> <p>Si se trata de una desconcentración mínima, el jerarca pierde las facultades de avocación y revisión con respecto al órgano desconcentrado.</p> <p>Si se trata de desconcentración máxima, además de lo anterior se encuentra sustraído de las instrucciones del jerarca.</p>	<p>Ser una instancia fuera de la estructura orgánica del MEP le brindaría mayor independencia funcional en su rol evaluador.</p> <p>Existen formas de estructurar su independencia (por ejemplo, institución autónoma, semiautónoma o ente público no estatal).</p>
Presupuesto	<p>Los recursos humanos y presupuestarios seguirían dependiendo del presupuesto del Ministerio, excepto si se le otorgara personalidad jurídica instrumental, lo cual requeriría de ley y estaría sometido a las reglas de la ley 9524, Ley de Fortalecimiento del control presupuestario de los órganos desconcentrados del Gobierno Central.</p>	<p>Al ser una instancia independiente del MEP, su presupuesto no estaría supeditado al Ministerio, sino que sería formulado por el ente y aprobado por la Contraloría General de la República.</p>
Flexibilidad para contratación de personal	<p>No podrían variarse las condiciones de contratación de personal que mantiene el MEP.</p> <p>Para la contratación de bienes y servicios (incluidos servicios profesionales) deberán seguirse las reglas previstas en la Ley General de Contratación Pública.</p>	<p>Su creación requiere necesariamente de una ley, en la cual podrían establecerse condiciones especiales de contratación temporal o permanente de personal según necesidades internas, aunque respetando los principios básicos del empleo público.</p>
Comunicación y articulación entre las partes	<p>Al ser un órgano de desconcentración dentro del Ministerio, la comunicación con otras dependencias se facilitarían. Asimismo, al existir un superior jerárquico común, cualquier discrepancia podría resolverse por éste.</p>	<p>La ley deberá establecer mecanismos de comunicación y articulación entre las diferentes partes, especialmente entre la instancia evaluadora, el MEP y el CSE. En caso de instituciones autónomas, estas cuentan con Juntas Directivas (órganos colegiados) que son integrados por representantes de otras instituciones con interés en la materia que regulan.</p>
Competencia del CSE	<p>El CSE mantendría su competencia de dirección general, estableciendo lineamientos y guías de evaluación que deben ser ejecutadas por dicho órgano.</p>	<p>El CSE mantendría su competencia de dirección general, estableciendo lineamientos y guías de evaluación que deben ser ejecutadas por esta instancia.</p>

Sistema de medición

Definición del problema de política pública a resolver

Los instrumentos de medición actuales carecen de estabilidad y no se articulan con los aprendizajes esperados dictaminados en la política curricular. Pese a la trayectoria nacional en macroevaluaciones, cada cierto tiempo se modifican las características o se suspenden pruebas sin que estas decisiones se basen en criterios técnicos. Lo anterior ocasiona debilidades para medir los aprendizajes dictaminados en la política curricular. Además, dificulta el seguimiento de logros y trayectorias educativas, la asimilación y aprovechamiento de los resultados para la mejora educativa, la provisión de información para la rendición de cuentas del sistema y también la acumulación de capacidades técnicas, tanto a nivel del ente evaluador como de los usuarios de la información,

de los análisis y de las investigaciones que se generan a partir de los procesos evaluativos.

Consideraciones propositivas

Resulta muy difícil prosperar cuando las reglas del juego cambian con frecuencia. Las personas especialistas concuerdan en que el sistema de medición debe contar con planificación a mediano y largo plazo, pues no es sino hasta esos horizontes temporales cuando se empiezan a percibir los impactos de la evaluación. Como se mencionó en el apartado anterior, son diversos los actores que intervienen e importantes los esfuerzos de articulación requeridos para implementar mejoras.

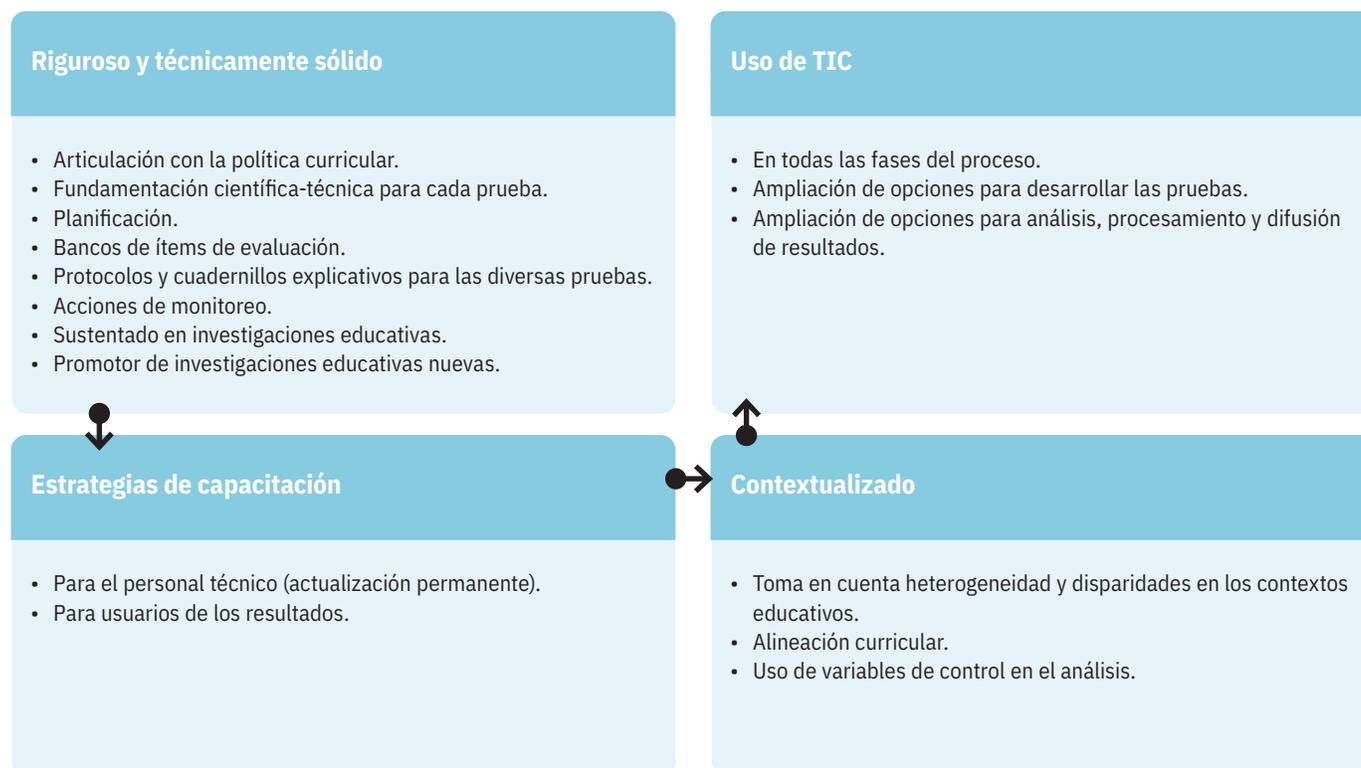
De las consideraciones anteriores no se deriva que las pruebas deban ser totalmente rígidas y que no puedan ajustarse en el tiempo, ya sea para acomodar las distintas necesidades a las que responde

la evaluación, introducir mejoras a los procesos existentes o bien para incorporar innovaciones metodológicas o tecnológicas. Por ejemplo, los programas de estudio de las diversas asignaturas se modifican cada cierto tiempo y las pruebas deben contemplar estos cambios.

Por lo tanto, es necesario que exista flexibilidad para evolucionar y que los cambios se basen en atender de mejor manera los objetivos y necesidades que se establecen en el marco teórico y normativo. Para lograrlo, se pueden considerar alternativas que involucren la sustentación de los cambios con criterios técnicos, la documentación de protocolos que incluyan los requisitos necesarios para llevar a cabo las modificaciones, así como sus implicaciones en la implementación y uso de los resultados. Es fundamental que estos cambios se comuniquen adecuadamente en los distintos niveles del sistema (figura 3.16)

Figura 3.16

Características claves del sistema de medición



La rigurosidad y solidez técnica se asienta en la fundamentación técnica-científica para cada una de las pruebas desarrolladas, sean estas sumativas (censales o muestrales), formativas, adaptativas, autoevaluaciones o coevaluaciones. Esta fundamentación científica se basa en metodologías y teorías probadas (recuadro 3.10). La planificación considera todas las etapas del proceso evaluativo: el diseño de pruebas, instrumentos y protocolos, la validación de la información generada, la captura y procesamiento de los datos, su análisis, la generación y difusión de reportes de resultados para distintos actores del sistema educativo, la revisión de los procesos de evaluación (innovación y mejora), la promoción de la reflexión educativa en diversos niveles y el fortalecimiento de las capacidades técnicas y de los usuarios. Contar con bancos de ítems de evaluación probados es relevante para alimentar diversos tipos de evaluaciones, como por ejemplo las formativas adaptativas o las autoevaluaciones.

Los protocolos y cuadernillos son importantes para docentes y centros educativos (quienes a menudo tienen a su cargo la aplicación de las distintas pruebas) pues detallan paso a paso los objetivos de la evaluación, así como las etapas y características de la implementación y el uso de sus resultados. Sobre este tema, Lasida (2023) recomienda que los protocolos estén asociados a las herramientas de evaluación y no a los puestos de trabajo de quienes las implementan.

En el caso de las evaluaciones formativas adaptativas, el ente evaluador puede colaborar en el desarrollo de instrumentos mediante la inclusión de una mirada disciplinar y didáctica en el diseño de actividades y recursos evaluativos. Para ello, puede facilitar el acceso al banco de ítems, por lo menos a aquellos que están accesibles (Lasida, 2023). También puede apoyar en la elaboración de lineamientos generales para su aplicación o incluso en la propia aplicación.

Por otra parte, es necesario clarificar la interrelación entre estas evaluaciones a nivel micro y las pruebas estandarizadas nacionales e internacionales, ya que las primeras brindan una oportunidad para

Recuadro 3.10

Metodologías empleadas en sistemas de evaluación latinoamericanos

Sin pretender un recuento exhaustivo, se recopilan algunas de las teorías y metodologías mencionadas por los expertos latinoamericanos consultados.

Ospina y Angulo (2023) y Medeiros (2023) definen la Teoría de Respuesta al ítem para la construcción de escalas de logro y niveles de desempeño.

En el caso colombiano (Ospina y Angulo, 2023) también se emplea el Diseño Universal de Evaluación que se sustenta en varios principios como la evaluación inclusiva, la precisión en la definición del constructo, el empleo de ítems no sesgados, la susceptibilidad para acomodar el contexto de manera que se garantice el acceso a diversas poblaciones, el uso de instrucciones y procedimientos sencillos,

claros e intuitivos, maximización de la lecturabilidad y comprensión y máxima legibilidad.

Miranda (2023) relata la adopción de Mapas de Progreso con base en el apego (o no) a los estándares e indicadores educativos con el fin de evidenciar logros también a nivel de las aulas. Del mismo modo, en Perú se han introducido los Tests Informáticos Adaptativos o TEI. Además, se utiliza la metodología *Bookmark* para realización de talleres “de puntos de corte” orientados al análisis de la dificultad de los ítems de evaluación.

Fuente: Vargas, 2023 con base en Medeiros, 2023; Ospina y Angulo, 2023; Lasida, 2023; Miranda, 2023 y Miranda y Monroy, 2023.

socializar la importancia de la evaluación. Las evaluaciones formativas y las autoevaluaciones pueden utilizarse en los análisis generales del sistema educativo para complementar los análisis de los resultados de pruebas estandarizadas censales o muestrales. El caso uruguayo constituye un ejemplo de interacción entre los distintos esfuerzos de evaluación que se realizan y cómo estos pueden llegar a repercutir a nivel de las prácticas pedagógicas en el aula (Lasida, 2023; Rodino, 2023a y Miranda, 2023).

El monitoreo y seguimiento de la ejecución de los procesos evaluativos es fundamental para definir rutas de mejora en el sistema de medición. En este proceso, la experiencia de varios países ha implicado la elaboración conjunta entre personas académicas, especialistas, centros que desarrollan pruebas a nivel internacional (como la OCDE con las pruebas PISA o IESALC con las pruebas ERCE) y/o entes evaluadores nacionales de probada trayectoria, como en el caso

de Australia (Miranda, 2023; Ospina y Angulo, 2023; Medeiros, 2023).

Una actividad fundamental para asegurar la solidez técnica y científica del sistema de medición son los procesos investigativos, tanto los que se basan en el proceso de evaluación en sí, como los que surgen a partir de los resultados evaluativos. Por ejemplo, puede indagarse sobre los factores asociados a los niveles de desempeño de la evaluación, sobre el diseño de diagnósticos de situación o sobre el cumplimiento de los objetivos del sistema, entre muchas otras posibilidades. Resulta indispensable contar con lineamientos, protocolos y productos de las investigaciones que permitan obtener conocimientos pertinentes para la mejora del proceso evaluativo y la calidad de la educación y para lograr que responda a los objetivos planteados en el marco de referencia.

Por su parte, las **estrategias de capacitación** son de relevancia tanto para el personal interno del ente evaluador

como para los potenciales usuarios de la información que se genera producto de los procesos evaluativos. Esta capacitación resulta fundamental para trabajar aspectos relacionados con cada una de las fases descritas en el apartado sobre la planificación.

Las personas expertas coinciden acerca de la importancia de tomar en cuenta las variables contextuales para las diferentes evaluaciones. Existe diversidad en las poblaciones evaluadas en cuanto a condición socioeconómica, acceso a diversos recursos educativos y tecnológicos, bagaje cultural y étnico, raza, género, idioma o estatus migratorio, entre otros criterios. Asimismo, se cuenta con distintas modalidades educativas, cada una con diferentes características y programas de estudio. Cada cohorte educativa en los centros puede requerir adecuaciones curriculares específicas, pues ningún grupo es igual a otro. Por su parte, escuelas y colegios tienen formas de operación diferentes y se encuentran inmersos en entornos que presentan climas de convivencia, condiciones, dotaciones y exposición a riesgos que le son propios.

Por ello, durante el análisis de resultados es virtual considerar estos factores, ya que debe definirse claramente cuáles de estos deben trabajarse como variables de control que permitan hacer comparaciones más justas a lo interno del sistema educativo. En los análisis también se pueden incluir estadísticas e información de bases de datos provenientes de otras fuentes que ayuden a complementar y mejorar la interpretación de los resultados (como es el caso de los datos censales o las Encuestas de Hogares).

Por último, una recomendación generalizada tiene que ver con el **uso de TIC** en todas las fases de los procesos evaluativos (diseño, aplicación, captura de datos, procesamiento, y difusión), lo cual puede incidir en la reducción de algunos costos del proceso (particularmente en la aplicación). Sin embargo, de igual forma, se deben contemplar los contextos desiguales en términos de acceso a equipos electrónicos, software y conectividad.

El uso de las TIC permite un aumento de las posibilidades para desarrollar evaluaciones (por ejemplo, mediante

diferentes dispositivos electrónicos). También permite incluir a personas con discapacidades o necesidades lingüísticas especiales y brinda la oportunidad de diseñar pruebas innovadoras mediante estímulos audiovisuales, entornos interactivos y algoritmos de inteligencia artificial. Además, se pueden desarrollar pruebas adaptativas que consideren el desempeño individual de cada estudiante. Finalmente, disponer de mayores capacidades digitales permite enriquecer los análisis a través de la ciencia de datos y reducir los tiempos de entrega de los resultados. Esto requiere, sin embargo, el fortalecimiento de las capacidades digitales de los equipos técnicos, de los implementadores y de los usuarios de la información.

Nota especial: el caso del diseño y construcción de una Prueba Nacional de Matemáticas articulada con el programa de estudios

Un ejemplo concreto de aplicación en torno a la necesidad de contar con sistemas robustos de medición alineados con el currículo y técnicamente bien fundamentados lo proporciona la Prueba Nacional de Matemáticas. Poveda et al. (2023) propone una ruta para diseñar una prueba estandarizada articulada con las propuestas del programa de estudios, la cual comprende cinco ámbitos de acción que se detallan en las siguientes líneas.

El primer ámbito implica la definición del qué y para qué evaluar, indispensable para velar por el cumplimiento de las aspiraciones de la política educativa y curricular asumida por el ente ministerial. En el caso de la prueba de Matemática, esta definición debe corresponder con el constructo de “competencia matemática” definido en los programas de estudio como:

una capacidad del individuo para formular, emplear e interpretar las Matemáticas en una variedad de contextos. Incluye razonar matemáticamente y usar conceptos, procedimientos, hechos y herramientas para describir, explicar y predecir fenómenos. Ayuda a los individuos a



El uso de TIC en evaluación no significa trasladar la aplicación de las pruebas de lápiz y papel a medios electrónicos, sino más bien de considerar todas las capacidades que brindan los sistemas informatizados (Miranda y Monroy, 2023).

reconocer el papel de las Matemáticas en el mundo y hacer juicios bien fundados y decisiones necesarias para ciudadanos constructivos, comprometidos y reflexivos” (MEP, 2012).

Sus dimensiones se presentan en la figura 3.17.

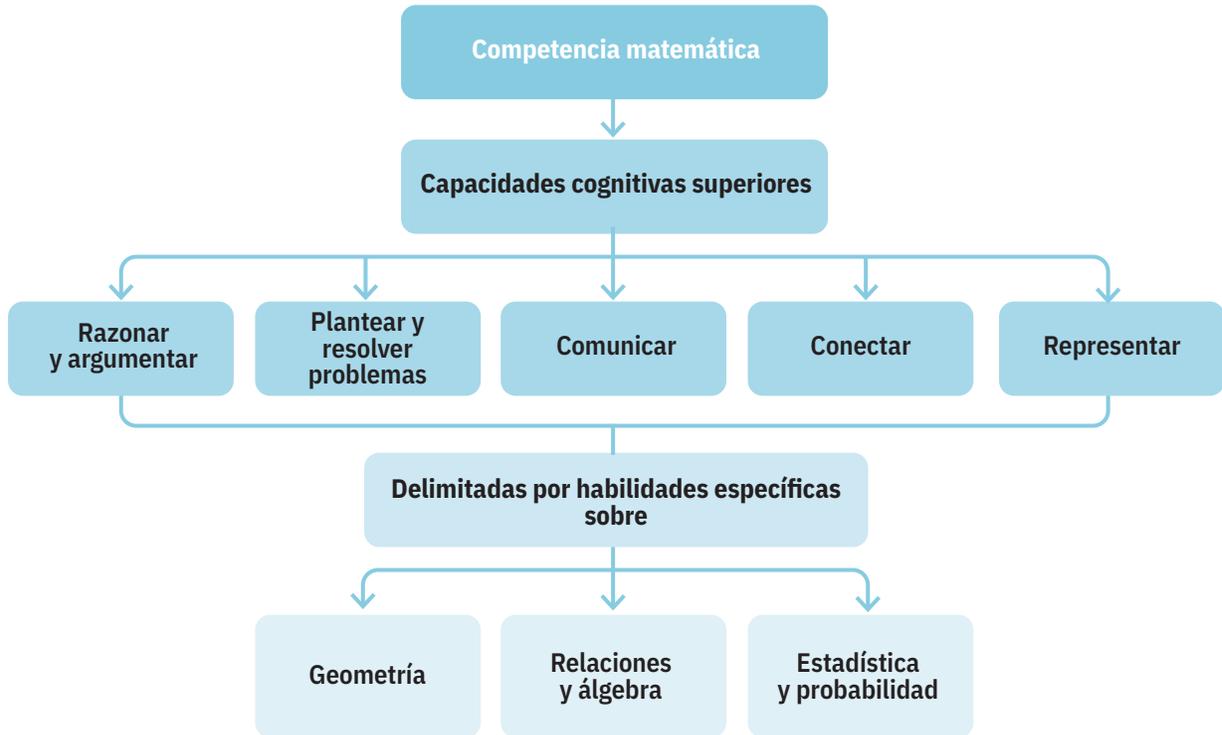
El segundo ámbito corresponde a implementar una **metodología especializada para la construcción de ítems**. La literatura señala que la elaboración de ítems pertinentes requiere de un equipo especializado en su confección que debe apegarse a diversas fases y pasos (figura 3.18).

El tercer ámbito tiene que ver con garantizar el ensamblaje de la prueba con criterios cualitativos y cuantitativos para juzgar adecuadamente los ítems. En este campo se proponen las siguientes acciones:

- Considerar criterios cuantitativos y cualitativos asociados a la valoración del nivel de complejidad de las tareas matemáticas desde la perspectiva que proponen los programas de estudio de matemáticas y su fundamentación teórica.
- Incluir diversos tipos de ítems²⁰ (en especial de respuesta construida) para evaluar procesos matemáticos como razonar, argumentar, plantear y resolver problemas o comunicar, conectar y representar los objetos matemáticos.
- El proceso de elegir “ítems ancla” y elaborar los cuadernillos debe considerar la distribución uniforme de ciertos componentes como el nivel de

Figura 3.17

Dimensiones para medir el constructo de competencia matemática en una prueba nacional estandarizada



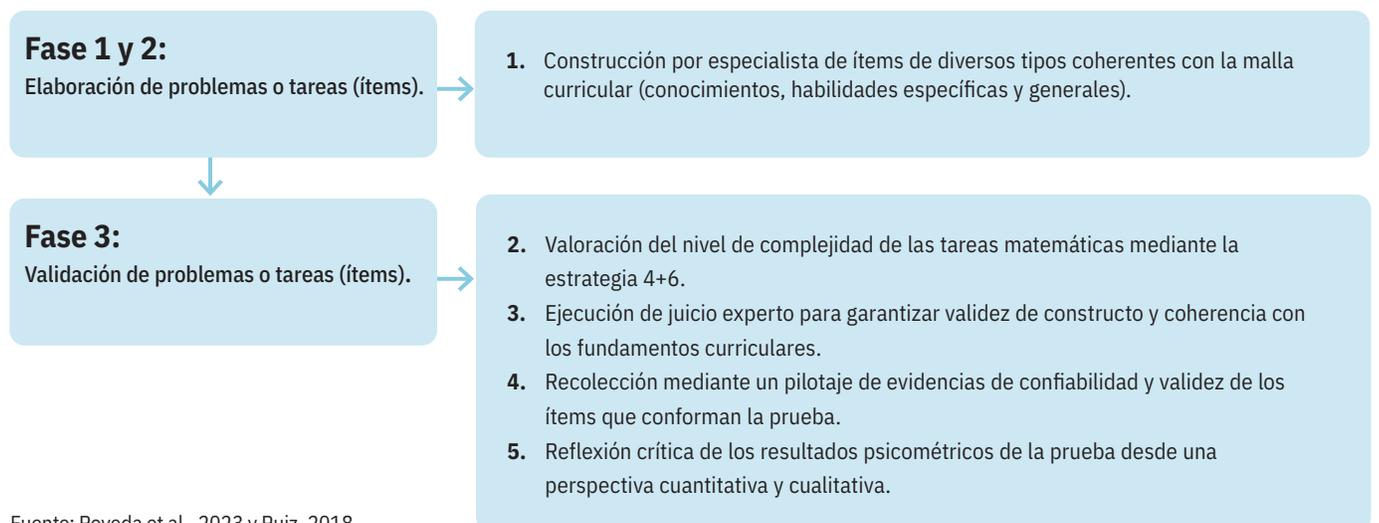
Fuente: Poveda et al., 2023.

Figura 3.18

Fases y pasos por seguir para confeccionar ítems pertinentes con el Programa de Estudios de Matemáticas

Fases

Pasos



Fuente: Poveda et al., 2023 y Ruiz, 2018.

complejidad (reproducción, conexión o reflexión) y la dificultad. Ruiz (2018) sugiere que los niveles de complejidad de las tareas matemáticas sean solo una guía para la construcción de ítems de prueba y plantea que estos deben emplearse en conjunto con las capacidades matemáticas superiores. Este autor recomienda una proporción de 25% de ítems de reproducción, 50% de conexión y 25% de reflexión en la prueba.

- Se debe tener extremo cuidado en la asignación del número de preguntas, su exigencia cognitiva y el tiempo correspondiente, debido a que una prueba con estas características requiere de una excelente calibración.

El cuarto ámbito es la aplicación de rigurosos protocolos para el análisis de resultados. Se propone implementar los protocolos establecidos por el ente ministerial y explorar la posibilidad de incluir recursos tecnológicos como plataformas de aprendizaje y herramientas digitales para el juzgamiento de respuestas construidas. El análisis de los resultados debe realizarse mediante procesos cualitativos y cuantitativos. Los resultados se deben discutir en función de la exigencia cognitiva de las tareas matemáticas y deben complementarse con datos psicométricos. Es necesario evitar la toma de decisiones basadas únicamente en los dos parámetros de medición. El propósito de la prueba (diagnóstica o de certificación) condiciona todo lo asociado a ella, incluso la presentación de los resultados y el uso que se da a estos.

Información y usos de resultados

Definición del problema de política pública a resolver

En Costa Rica la información evaluativa no se está utilizando para desarrollar acciones y políticas para el mejoramiento de la calidad educativa que vayan más allá de la simple comunicación de las deficiencias en los procesos educativos del estudiantado. Una de las principales debilidades que enfrentan los sistemas

nacionales de evaluación analizados en la elaboración de este capítulo especial tiene que ver con la deficiente devolución de los resultados a los distintos actores del sistema educativo, tanto a las autoridades gubernamentales, como a la sociedad en su conjunto.

Consideraciones propositivas

El consenso de los autores consultados es que de nada sirve contar con evaluaciones de gran calidad técnica que generen información pertinente para satisfacer los objetivos del marco de referencia definido si esa información no llega en tiempo y forma a quienes tienen responsabilidades para impulsar las mejoras necesarias.

En el país la norma ha sido difundir la “mala noticia” de los pobres resultados que obtienen los estudiantes en las distintas pruebas estandarizadas nacionales e internacionales. Sin embargo, más allá de eso, no se canaliza la información, el análisis y los estudios derivados de esa evaluación hacia los tomadores de decisiones. Las principales limitaciones se dan, como en la mayoría de los países, entre los resultados agregados a nivel nacional y las personas docentes que deberían implementar cambios pedagógicos en las aulas.

La creación de vasos comunicantes entre los distintos actores que inciden sobre las evaluaciones y la promoción de la reflexión-acción educativa son fundamentales para cimentar la legitimidad del sistema evaluativo. Desde esta óptica, resulta necesario avanzar del levantamiento de la información hacia su uso para reorientar las prácticas pedagógicas. El estudio realizado por Francis (2023b) recopiló las voces de los actores educativos en cuanto a las transformaciones requeridas para que el personal docente considere la información desprendida de la evaluación como insumos para la toma de sus decisiones y actuaciones pedagógicas. Estas transformaciones se deben realizar en tres áreas: generación de informes especializados para cada audiencia, producción de información para orientar la toma de decisiones pedagógicas e implementación en distintos formatos tecnológicos (figura 3.19).

Francis (2023b) señala que el aprovechamiento de la información anterior requiere que la gobernanza provea las siguientes acciones específicas: en primer lugar, debe desarrollar procesos de actualización profesional para docentes y directivos enfocados en mejorar la comprensión macroevaluativa, su diseño e implementación, incluyendo especificidades como el diseño de pruebas estandarizadas, tipos de ítems y la interpretación de resultados técnicos. Además, se busca clarificar el papel de la evaluación y cómo puede articularse en los planes de mejora con los programas de estudio.

En segundo lugar, debe impulsar estrategias de seguimiento sostenido en el centro educativo mediante la atribución de un papel más de acompañamiento por parte del Comité de Evaluación y contar con espacios institucionales y flexibles para crear rutas atinentes a las características de cada institución, circuito y contexto socioeconómico. En tercer lugar, es necesario revisar la gestión institucional del MEP, puesto que los cambios informativos abruptos sobre el proceso de macroevaluación generan inestabilidad y limitan la capacidad de tomar decisiones. Finalmente, se sugiere revisar en los programas de formación de las universidades cómo se orienta al docente en formación para reconocerse como actor responsable de la mediación docente y, por tanto, con capacidad para analizar los distintos insumos que le permiten transformar sus prácticas pedagógicas.

Los resultados de las evaluaciones pueden incidir en la definición de políticas públicas, como por ejemplo en la focalización de intervenciones de las autoridades educativas, la definición de políticas de supervisión de centros educativos, la construcción de mapas de atención prioritaria, el diseño de estrategias de acompañamiento pedagógico a los docentes o el diseño curricular. En algunos países, se han utilizado para focalizar recursos económicos (por ejemplo, inversiones en infraestructura, equipamiento o tecnología, concesión de incentivos monetarios para centros educativos, directores o docentes o el otorgamiento de subsidios o becas a estudiantes y docentes, como sucede en Chile (Medeiros, 2023)

Figura 3.19

Recomendaciones para generar información de la macroevaluación que promueva prácticas pedagógicas efectivas, según actores educativos^{a/} consultados

Generar informes especializados para cada audiencia	Crear contenido en los informes para la toma de decisiones pedagógicas	Proveer distintos formatos de información
<p>Involucrar al personal docente en el proceso de macroevaluación: en la mediación, definición de tabla de especificaciones y construcción de ítems.</p> <p>Considerar a los y las docentes de todos los niveles como audiencias específicas de la información para fomentar una cultura de mejora continua en todos los años escolares o colegiales.</p> <p>Impulsar jornadas de análisis por asignaturas, enfocadas en la mediación pedagógica y no solo en la prueba, para promover la formación entre pares y consolidar redes de apoyo con directores y asesores.</p> <p>Identificar líderes pedagógicos para interpretar los informes de evaluación y trabajar con el personal docente a través del fomento de la reflexión y del cambio en las prácticas de aula.</p> <p>Fomentar una actitud autocrítica y empoderada en el análisis y reflexión de los resultados al convertir los informes en una herramienta de mejora de la práctica docente y no en una tarea adicional.</p>	<p>Incluir los logros y áreas de mejora de los estudiantes: identificar fortalezas y debilidades y definir los niveles de habilidad en los que se ubica el estudiantado.</p> <p>Redactar con lenguaje claro y accesible: incluir descripciones narrativas para facilitar la comprensión de gráficas y tablas y establecer conclusiones claras en un lenguaje sencillo, no técnico.</p> <p>Incluir orientaciones didácticas con recomendaciones específicas para mejorar las habilidades y conocimientos no alcanzados por los estudiantes, tales como ítems más difíciles, estrategias cognitivas, dificultades detectadas, observaciones conceptuales y propuestas metodológicas para superar las dificultades.</p> <p>Contextualizar los resultados según los recursos, perfiles docentes, problemáticas sociales y compromiso de las familias, con el fin de que se adapten a las diferentes realidades y promuevan acciones específicas.</p>	<p>Inclusión de TIC para contar con disponibilidad de información a tiempo y articulada con los programas de estudio.</p> <p>Utilizar herramientas de inteligencia artificial para personalizar el uso de los resultados educativos.</p>

a/La consulta incluyó entrevistas grupales e individuales a 25 docentes de colegios y escuelas en zonas urbanas y rurales, un supervisor y dos directores de instituciones educativas.

Fuente: Barquero, 2023 con base en Francis, 2023b.

y Perú (Miranda, 2023). Asimismo, las evaluaciones brindan información relevante para evaluar los impactos de las políticas o medidas adoptadas, lo que permite, a su vez, revisar, complementar, o avanzar en el diseño de nuevas políticas.

Estas acciones de información y orientación de los usos de los resultados evaluativos ayudan a que estos dejen de verse como una “carga”, sobre todo entre quienes tienen la responsabilidad de su financiamiento e implementación (Medeiros, 2023).



Es necesario evitar la deslegitimación de la evaluación por el mal uso de los resultados (Lasida, 2023). Un ejemplo de mal uso de resultados es la elaboración de rankings descontextualizados que propician fuga de matrícula de algunos centros (Medeiros, 2023; Miranda, 2023) o que afectan

la permanencia de docentes si no se garantizan procesos de capacitación y apoyos para que puedan trasladar sus resultados de la evaluación a su quehacer en las aulas (Miranda y Monroy, 2023).

Los esfuerzos comunicativos y el soporte a los usos de la información requieren de un cuidadoso proceso de planificación. Es necesario diseñar e implementar una estrategia en la que estén claramente definidos los objetivos acerca de lo que se busca comunicar, los temas o aspectos que se comunicarán, así como la forma, los formatos y los medios que se emplearán para la difusión (figura 3.20).

Ciertamente, la información que se presente debe estar bien hilada y argumentada, pero además debe ser defendible y metodológicamente sólida. De igual forma, es vital que las comunicaciones se adapten al lenguaje, a las necesidades y a los diferentes formatos que las personas usuarias emplearán para procesarlas.

Esto implica que se requerirán diversas versiones de los mismos resultados, pero con énfasis u ópticas diferentes.

La idea central aquí es que las acciones deben orientarse más a divulgar información y conocimientos de manera responsable y con argumentada y menos a diseminar datos sin contexto. Es posible que en algunos reportes deba prescindirse del uso de lenguajes excesivamente técnicos para favorecer una narrativa que ayude a la comprensión de los resultados y de su utilidad. Asimismo, en algunos casos será necesario enfocarse más en los procesos que en los propios resultados.

Finalmente, una consulta a expertos nacionales arrojó recomendaciones para



La capacidad técnica de los usuarios puede limitar el uso de los resultados, por lo que es importante contemplar mecanismos que apoyen su desarrollo (Ospina y Angulo, 2023).

construir un sistema de macroevaluación en el país que se detallan en el recuadro 3.11 y proveen ideas para adaptarse a las realidades concretas del país, así como sugerencias planteadas por las personas expertas de otros países de América Latina y la revisión de buenas prácticas internacionales.

Figura 3.20

Algunos elementos para elaborar una estrategia de comunicación y orientación de usos de los resultados evaluativos

Objetivos	Aspectos a comunicar	Medios
<p>Divulgar y promover el valor social de la evaluación y sus resultados para la mejora educativa.</p> <p>Fomentar la cultura de evaluación en todos los niveles del sistema educativo. Socializar los resultados.</p> <p>Socializar alternativas pedagógicas.</p> <p>Informar sobre el desarrollo de políticas públicas orientadas a mejorar la calidad de la educación en el país y su efectividad.</p> <p>Informar sobre el cumplimiento de los objetivos del sistema.</p>	<p>Resultados de las pruebas (incluyendo factores asociados).</p> <p>Hallazgos de las investigaciones realizadas.</p> <p>Resultados del monitoreo y seguimiento (trayectorias educativas, procesos evaluativos, efectividad de medidas implementadas y cumplimiento de objetivos).</p> <p>Áreas de mejora identificadas.</p> <p>Uso pedagógico de resultados. Informar el desarrollo de políticas públicas orientadas a mejorar la calidad de la educación en el país.</p> <p>Innovaciones evaluativas (herramientas, procesos, etc.).</p> <p>Ítems liberados para la implementación del currículum.</p>	<p>Informes generales.</p> <p>Reportes específicos (infografías, videos y visores).</p> <p>Bases de datos (guardando los criterios de confidencialidad establecidos).</p> <p>Medios de comunicación masiva (televisión, radio o prensa).</p> <p>Página web institucional.</p> <p>Redes sociales (con debidos controles de información).</p> <p>Guías de uso e interpretación de datos.</p> <p>Orientaciones didácticas específicas para los docentes en las aulas.</p> <p>Talleres de información de capacidades.</p> <p>Comisiones y consejos en los que participa el ente evaluador.</p>

Recuadro 3.11

Recomendaciones de personas expertas nacionales para construir un sistema de macroevaluación en Costa Rica

Se realizó un taller de consulta con 6 personas expertas nacionales para establecer recomendaciones sobre las características que debe tener un sistema de macroevaluación de aprendizajes en Costa Rica. Estas se organizaron en cada uno de los componentes claves del sistema.

Sobre el marco de referencia:

- Precisar el qué y para qué de la evaluación estandarizada, así como sus alcances. La única finalidad no debe ser la evaluación de estudiantes.
- Generar estándares de evaluación para las competencias estudiantiles, docentes y de centros educativos en armonía con el currículo.

Sobre la gobernanza:

- Crear una agencia autónoma del MEP encargada de gobernanza técnica de los procesos macroevaluativos con las siguientes características: nítida separación de funciones; buena coordinación entre los entes políticos y el técnico; instancias pequeñas de dirección; autonomía técnica y potestad de hacer contrataciones de personal, establecimiento de convenios y búsqueda de los equilibrios en la aplicación de los instrumentos (las pruebas). Su creación contempla las siguientes alternativas: modificar la Constitución Política para poder existir con independencia del MEP y el CSE, concebirse como un órgano adscrito al CSE pero con autonomía técnica o bien adscribirse a un convenio entre el CSE y otras instituciones donde hay más experiencia (por ejemplo las universidades públicas).

- Sugerir la posible intervención de la Asamblea Legislativa para dar apoyo al impulso al ente independiente cuya consolidación va a requerir tiempo y recursos.
- Realizar consultas legales para considerar las posibilidades que existen y evitar que la gobernanza técnica no pueda gestarse.
- Fortalecer la coordinación y cooperación de los niveles y entidades que tiene que ver con la evaluación: MEP, universidades, colegios profesionales, INA, etc.

Sobre el sistema de medición:

- Incorporar a la evaluación docente y de centros educativos, no solo la evaluación a estudiantes.
- Garantizar absoluta correspondencia y alineación con los programas de estudio.
- Incluir instrumentos de contexto que permitan analizar diferencias entre las poblaciones y sugerir apoyos y medidas afirmativas para los grupos más vulnerables.
- Incorporar procesos para integrar las pruebas nacionales con las estandarizadas en las que el país participa.
- Emplear tecnologías sin descuidar los contextos desiguales en los que está actualmente inmersa la población estudiantil.
- Aplicar ajustes razonables en pruebas estandarizadas.
- Ensamblar pruebas a partir de bancos de ítems debidamente validados a nivel estadístico y psicométrico. Se requiere la aplicación de pruebas piloto.

- Incluir ítems de respuesta abierta, ejecución o desempeño para evaluar procesos de razonamiento.
- Realizar capacitaciones a las diferentes audiencias sobre los insumos que se generen de la macroevaluación para aprovechar el potencial de la información.
- Construir descripciones de los desempeños que den respuesta a qué puede hacer o no la persona examinada de acuerdo con el puntaje obtenido.
- Atender las inequidades de alfabetización tecnológica. Implica reconocer y contribuir a minimizar los problemas de validez y justicia asociados a esta temática.
- Valorar la aplicación de pruebas de distintos tipos según el objetivo y alcance: diagnóstica, de certificación, de altas o bajas consecuencias, según corresponda.
- Incluir instrumentos de medición actitudinales y socioemocionales.
- Considerar la evaluación formativa y sumativa.

Sobre la difusión y usos de la información:

- Acoger las buenas prácticas internacionales analizadas en este Informe.
- Integrar a los actores educativos (docentes, asesores y directivos) como miembros activos del proceso. Conocer sus necesidades de uso de información que permitan realizar las transformaciones pedagógicas necesarias.

Fuente: Barquero, 2023 a partir de la consulta realizada a expertos nacionales.



Para más información sobre

ASPECTOS CLAVE PARA UN SISTEMA NACIONAL DE MACROEVALUACIÓN DESDE LA EXPERIENCIA INTERNACIONAL

véase Medeiros, 2023; Ospina y Angulo, 2023; Miranda, 2023; Lasida, 2023 y Miranda y Monroy, 2023, en www.estadonacion.or.cr

Investigadores principales: María Victoria Angulo, Edwin Chaves, Susan Francis, Javier Lasida, María Paz Medeiros, Liliana Miranda, Francisco Miranda, Deborah Monroy, Eiliana Montero, Mónica Ospina, Ricardo Poveda, Ana María Rodino, Marianela Zumbado.

Insumos: Análisis de la gobernanza y el proceso de macro-evaluación en Costa Rica: evolución, alcances y desafíos para promover la mejora del sistema educativo, elaborado Susan Francis. Requerimientos de fondo y forma de los informes de pruebas nacionales para la macroevaluación como mecanismos para retroalimentar la mediación docente, elaborado por Susan Francis. Criterios y ruta para el diseño de pruebas nacionales consistentes con los programas de estudio de Matemáticas, elaborado por Ricardo Poveda, Edwin Chaves y Marianela Zumbado. Buenas prácticas nacionales e internacionales de macro evaluación. Lecciones para el diseño de una hoja de ruta de mejora del sistema educativo costarricense, elaborado por Ana María Rodino. Aportes para una hoja de ruta hacia el mejoramiento de la macro evaluación en la educación costarricense, elaborado por Ana María Rodino. Aspectos claves a considerar en el desarrollo de un sistema nacional público de macroevaluación de los aprendizajes de los estudiantes: recomendaciones desde el caso uruguayo, elaborado por Javier Lasida. Aspectos claves a considerar en el desarrollo de un sistema nacional público de macroevaluación de los aprendizajes de los estudiantes: recomendaciones desde el

caso peruano, elaborado por Liliana Miranda. Aspectos claves a considerar en el desarrollo de un sistema nacional público de macroevaluación de los aprendizajes de los estudiantes: recomendaciones desde el caso colombiano, elaborado por Mónica Ospina y María Victoria Angulo. Aspectos claves a considerar en el desarrollo de un sistema nacional público de macroevaluación de los aprendizajes de los estudiantes: recomendaciones desde el caso mexicano, elaborado por Deborah Monroy y Francisco Miranda. Aspectos claves a considerar en el desarrollo de un sistema nacional público de macroevaluación de los aprendizajes de los estudiantes: recomendaciones desde el caso chileno, elaborado por María Paz Medeiros. La importancia de los estándares educativos para evaluar el progreso de la educación costarricense, elaborado por Eiliana Montero.

Borrador de capítulo: Katherine Barquero y Jimena Vargas.

Coordinación: Katherine Barquero.

Asistente de investigación: Nicole Cruz.

Edición técnica: Katherine Barquero, Jorge Vargas Cullrell, Isabel Román.

Actualización, procesamiento y visualización de datos: Katherine Barquero.

Lectores críticos:

Clotilde Fonseca (CC), Steffan Gómez-Campos (PEN), Alberto Mora (PEN).

Por la participación como grupo de expertos,

se agradece a: Eiliana Montero (UCR), Vanessa Smith (UCR), Ángel Ruiz(CC), Andrea González (Fundación Yamuni), Luis Salazar (Fundación Yamuni), Marvin Carvajal (Fundación Yamuni), Ana María Rodino (CC), Susan Francis (UCR), María Victoria Angulo (Colombia), Javier Lasida (Uruguay), María Paz Medeiros (Chile), Francisco Miranda (México), Deborah Monroy (México), Javier Lasida (Uruguay), María Victoria Angulo (Colombia), Mónica Ospina (Colombia) y Liliana Miranda (Perú), Paulo Santiago (OCDE).

Por el apoyo en la coordinación de las entrevistas y grupos focales con docentes, directores y supervisores educativos, se agradece a:

Jorge Quesada (Colypro).

Revisión y corrección de cifras:

Katherine Barquero.

Corrección de estilo y edición de textos:

Diego Ugalde.

Diseño y diagramación:

Erick Valdelomar | Insignia ng.

Notas

1 Los resultados de las Pruebas Nacionales se limitaron al uso de los datos estadísticos sin apoyos didácticos, curriculares y metodológicos para facilitar su interpretación y su uso pedagógico para mejorar los procesos educativos.

2 Aún en este restringido ámbito, no han considerado la existencia de diversas posibilidades de implementación, ya sea a todo el estudiantado o a una muestra representativa de este en un determinado nivel escolar sobre el currículo de ciertas asignaturas.

3 Rodino (2023a) señala que la evaluación diagnóstica puede aportar información para identificar estudiantes en riesgo de fracaso al evidenciar las fuentes de sus dificultades de aprendizaje, lo cual implicaría el diseño de una intervención pedagógica complementaria o remedial.

4 Se realizaba una primera evaluación en quinto año de la primaria y décimo año de la secundaria y una segunda evaluación en sexto y undécimo o duodécimo con el fin de promover mejoras en los desempeños estudiantiles.

5 Los actores corresponden a los siguientes: los educandos y sus familias; el personal docente, los y las directores y otras personas funcionarias de los centros; el centro educativo como tal (los estudios basados en evidencia muestran que es un actor en sí mismo, como conjunto, más de allá de los actores mencionados antes, que lo integran); los roles intermedios del sistema educativo; las autoridades del gobierno de la educación, entre las que se destacan el Ministro y el Consejo Superior de Educación y entre las que debe considerarse a otros roles tanto del Poder Ejecutivo como del Legislativo; las personas académicas y técnicas, que producen conocimiento sobre educación y las organizaciones que los nuclean, como por ejemplo las universidades o los organismos internacionales; y la opinión de pública, que se expresa como tal (por ejemplo, en encuestas sobre educación). La opinión pública se conforma por distintos actores formadores de opinión en educación, entre los que se destacan los periodistas que abordan el tema educativo, los partidos políticos, especialmente a través de los capítulos educativos de los programas que proponen a la ciudadanía y también los sindicatos y otras organizaciones de la sociedad civil que actúan en relación con la educación.

6 A través del decreto Ejecutivo N°40862-MEP del 12 de enero de 2018 “Reglamento de evaluación de los aprendizajes” y sus reformas CSE-SG-0057-2023 del 31 de enero de 2023.

7 Se desarrollaron entrevistas individuales y grupales con personal docente de distintas zonas del país, tanto de colegios y escuelas urbanas, como de zonas rurales, del Valle Central y de zonas costeras, así como con personal docente de colegios que atienden población indígena, población en edad adulta, adolescentes, y población infantil. En total, se realizaron cuatro entrevistas grupales y nueve individuales, en las cuales participó un total de veinticinco profesionales docentes. Adicionalmente, se entrevistó a un supervisor y dos directores de institución educativa.

8 De acuerdo con el artículo 81 del Reglamento de Evaluación de Aprendizajes, los objetivos de las pruebas nacionales son: “Contribuir a la formación integral de la persona estudiante; coadyuvar en la determinación de la promoción o la certificación de la persona estudiante; conocer los resultados de logro de los objetivos curriculares de la educación costarricense, basado en criterios técnicos de medición y evaluación, de modo que permitan mejorar la calidad del sistema educativo en todos sus niveles y procesos; incorporar con base en los resultados obtenidos por el estudiantado en las respectivas Pruebas Nacionales, según lo permita esta información, las estrategias y propuestas conducentes al mejoramiento cualitativo de los procesos de enseñanza y aprendizaje, en aquellas áreas donde el sistema educativo lo requiera; ofrecer al estudiantado un desafío académico, que contribuya a mejorar las posibilidades de éxito para su incorporación a los ciclos o niveles educativos inmediatos superiores; promover una actitud de superación académica en el personal docente, motivándolos para que aporten lo mejor de sus conocimientos en la búsqueda de un mayor y mejor aprendizaje de los educandos; fungir como un recurso adecuado para el proceso de evaluación y control del rendimiento escolar”.

9 De hecho, en los años 90 y posteriores se publicaron gran cantidad de artículos sobre esta temática en una de las revistas más conocidas de dicha asociación, llamada *Educational Measurement: Issues and Practice* (<https://www.researchgate.net/journal/Educational-Measurement-Issues-and-Practice-1745-3992>).

10 Además, era necesario implementar un principio básico: las pruebas (estandarizadas y de aula) deben derivarse y ser consistentes con los procesos de mediación pedagógica, y no al revés. Más aún, el enfoque sistémico de la transformación curricular “Educar para una ciudadanía” y su marco de referencia, la iniciativa internacional ATC21 (Assessment and Teaching for 21st Century), también plantean explícitamente que las pruebas estandarizadas y de aula deben ser congruentes y basarse en los procesos de mediación pedagógica (Fundación Omar Dengo, 2014; Consejo Superior de Educación, 2016; Ministerio de Educación Pública, 2015; Griffin et al., 2012).

11 Los reglamentos corresponden a: Reglamento de exámenes públicos para los Institutos Nacionales de segunda enseñanza y estudios normales país aprobado entre 1889 y 1893 a través del Decreto n° 4 de José J. Rodríguez, presidente constitucional de Costa Rica firma y al Reglamento General de Bachillerato de los Establecimientos de Educación Media Oficiales y Particulares promulgado en 1968.

12 Las pruebas diagnósticas fueron de carácter muestral y se aplicaron en sexto año.

13 Entre las prácticas regionales no se incluye el caso de México. Si bien ese país contó por 16 años con el Instituto Nacional de Evaluación Educativa como órgano constitucional autónomo a cargo de la evaluación nacional del desempeño estudiantil, de docentes y de centros educativos, luego de pugnas, protestas y un fuerte debate parlamentario, la institución desapareció en 2018. En su lugar, se creó la Comisión Nacional para la Mejora Continua de la Educación como cuerpo descentralizado, pero no autónomo, y la evaluación a nivel nacional pasó a definirse como diagnóstica, voluntaria, participativa y comunitaria. La Comisión está finalizando el desarrollo y pilotaje de un nuevo modelo de evaluación, cuyos resultados aún no se conocen.

14 Entre los países europeos que cuentan con agencias evaluadoras están Dinamarca, Eslovaquia, Eslovenia, Finlandia, Francia, Hungría, Irlanda, Irlanda del Norte, Islandia, Italia, Luxemburgo, Noruega, Países Bajos, Polonia, Portugal, República Checa y Suecia. Los demás países reseñados por Rodino (2023a) son Australia, Canadá, Corea del Sur y Nueva Zelanda.

15 Los estándares educativos son descripciones de lo que los estudiantes deben saber (estándares de conocimiento) y ser capaces de hacer (estándares de desempeño) en diferentes etapas del proceso de aprendizaje.

16 Costa Rica ha participado en los siguientes estudios de factores asociados: Serce 2006 (Segundo Estudio Regional Comparativo), Terce 2013 (Tercer Estudio Regional Comparativo) y ERCE 2019 (Cuarto Estudio Regional Comparativo) del Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación (Llece) de Unesco. Además, participó en PISA 2009, 2012, 2015 y 2018, de la OCDE.

17 Las pruebas PISA, que se aplican cada 3 años desde el 2000 en los países y territorios miembros y asociados a la OCDE también cuentan con una sección de factores asociados. Dadas las diferentes características de los sistemas educativos, la población evaluada no se define en función del grado escolar que cursa sino de su edad (abarca a estudiantes con edades entre los 15 años y 3 meses y los 16 años y 2 meses), para garantizar que los grupos sean equivalentes y los resultados de todos los países puedan ser comparables. PISA evalúa competencias de los estudiantes en Lectura, Matemática y Ciencias, priorizando una de estas áreas en cada aplicación consecutiva.

18 La Universidad de Costa Rica, a través del programa de Evaluación en Lenguas Extranjeras (Pelex) de la Escuela de Lenguas Modernas ha realizado alianzas estratégicas con organismos internacionales que trabajan con IA. Ha trabajado en conjunto para mejorar los algoritmos con evaluadores humanos especialistas de la UCR y así contar con una mayor precisión en las puntuaciones de la evaluación oral. Esta herramienta de evaluación lingüística apoyada con IA se puede utilizar con fines de diagnóstico o de ubicación, es fácil de usar, y los estudiantes pueden tomarlo en línea desde cualquier lugar y en cualquier momento. Proporciona retroalimentación inmediata al estudiante e indica áreas donde necesita mejorar, como la pronunciación, la fluidez o la gramática, lo cual permite ahorrar tiempo y recursos y proporcionar evaluaciones confiables y objetivas del desempeño lingüístico de la población estudiantil (Quesada et al., 2023).

19 También tiene la responsabilidad de elaborar un examen centralizado basado en competencias.

20 El uso de ítems de selección múltiple para discriminar entre procesos resolutorios y la exigencia de la respuesta construida favorecen la recolección de evidencias sobre el desempeño del estudiantado respecto a esas acciones cognitivas (Zumbado, 2019).