



**ESTADO
DE LA EDUCACIÓN**

Noveno Informe Estado de la Educación (2023)

Investigación

Buenas prácticas nacionales e internacionales de macro evaluación. Lecciones para el diseño de una hoja de ruta de mejora del sistema educativo costarricense

Investigadora:

Ana María Rodino Pierri

San José | 2023



378
R692b

Rodino Pierri, Ana María

Buenas prácticas nacionales e internacionales de macro evaluación : lecciones para el diseño de una hoja de ruta de mejora del sistema educativo costarricense / Ana María Rodino Pierri. -- Datos electrónicos (1 archivo : 631 kb). -- San José, C.R. : CONARE - PEN, 2023.

ISBN 978-9930-618-55-4

Formato PDF, 46 páginas.

Investigación para el Noveno Informe Estado de la Educación (2023)

1. SISTEMA EDUCATIVO. 2.GOBERNANZA INSTITUCIONAL. 3. EVALUACIÓN EDUCATIVA. 4. COSTA RICA. I. Título.



Índice

Reconocimientos	4
Descargo de responsabilidad	4
La macro-evaluación de los aprendizajes y sus funciones. Un aspecto de la educación costarricense que necesita atención.....	4
Algunas precisiones sobre evaluación educativa	8
Buenas prácticas nacionales e internacionales sobre factores centrales de la macro-evaluación	10
Gobernanza y organización institucional.....	10
Objetivos	18
Técnicas e instrumentos de medición.....	23
Difusión y uso de resultados.....	31
Conclusiones y recomendaciones	36
Sobre la gobernanza y organización institucional de la macro-evaluación.....	37
Sobre los objetivos de la macro-evaluación	38
Sobre las técnicas e instrumentos de medición de la macro-evaluación	39
Sobre la difusión y uso de resultados de la macro-evaluación	40
Referencias	41

Reconocimientos

Asistente de investigación: Nicole Cruz Ceciliano

Descargo de responsabilidad

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

La macro-evaluación de los aprendizajes y sus funciones. Un aspecto de la educación costarricense que necesita atención

La *macro-evaluación* de los aprendizajes es aquella de alcance nacional que se aplica a todos o a una muestra representativa de estudiantes de determinado nivel escolar sobre el currículo de ciertas asignaturas de su nivel, con el fin de medir su desempeño y logros educativos. Este esfuerzo, que es considerable para el sistema educativo en inversión de recursos profesionales, financieros y técnicos, puede responder a distintas preocupaciones de las autoridades nacionales y la ciudadanía. Como bien señalan Ravela, Arregui, Valverde, Wolfe, Ferrer, Martínez Rizo, Aylwin y Wolff (2008), puede surgir de su interés en la formación ciudadana de los jóvenes para fortalecer la democracia, o en la capacitación de la futura fuerza laboral para impulsar la economía del país, o en el desarrollo integral de las personas para participar en la moderna sociedad del conocimiento, o en la equidad social y la educación como camino para superar la pobreza, o en una combinación de todos estos intereses. Pero, en cualquier caso, siempre se parte de que la macro-evaluación es un instrumento que, bien utilizado, contribuye a mejorar la gestión del sistema, la calidad de la enseñanza, y la definición de políticas para beneficio de la educación nacional.

Debido a su alcance, la macro-evaluación emplea *pruebas estandarizadas*, es decir, uniformes para todos y cada uno de los estudiantes y centros educativos por evaluar. Estas pruebas son instrumentos construidos por especialistas en psicometría y rigurosamente validados. Se aplican siempre a un conjunto amplio de individuos y se califican siguiendo estrictos protocolos a efectos de garantizar que las puntuaciones sean comparables y que las inferencias que se hagan sobre las personas examinadas sean válidas. (Rodino, 2013; IV Informe Estado de la Educación, 2013; E: Santiago, 2022)

Las macro-evaluaciones estandarizadas –sobre las que volveremos en la sección siguiente para contrastarlas con las micro-evaluaciones, que no lo son (ver Cuadro 1)— hacen aportes muy valiosos al sistema educativo que, sin ser absolutos, superan los de otros tipos de evaluaciones. Siguiendo a Ravela et al. (2008, p.46-48), repasemos esos aportes y su fundamentación:

- *Ayudan a visualizar los resultados educativos del conjunto de estudiantes.* Los resultados del esfuerzo educativo no se ven a simple vista, de manera directa e inmediata. Los buenos docentes pueden apreciar si sus estudiantes aprenden o no, pero la heterogeneidad del cuerpo docente de un país no permite alcanzar una visión de conjunto por la simple suma de puntos de vista individuales. Esa visión de conjunto es la que persigue la macro-evaluación.
- *Entregan información sobre el real acceso al conocimiento y las capacidades que logran los alumnos, más allá de su número de años de estudio.* Hace algunas décadas se entendía que acceder al conocimiento y las capacidades cognitivas superiores equivalía al tiempo de permanencia en el sistema educativo, por eso se medían con indicadores del sistema como matrícula, cobertura, retención, etc. Esta equivalencia actualmente se resquebrajó, en parte porque al universalizarse el acceso al sistema, muchos estudiantes de sectores desfavorecidos llegan y permanecen en la escuela con muchas desventajas, y en parte porque el crecimiento del acceso a la profesión docente no fue siempre acompañado con garantías de calidad en la formación. Así, más años de educación no implican necesariamente que el estudiantado incorpore más conocimientos, actitudes y capacidades complejas. Las macro-evaluaciones ilustran mejor lo que ocurre en la realidad.
- *Ayudan a hacer visibles un conjunto de aspectos centrales de la labor educativa.* Entre otras cosas, muestran en qué medida los estudiantes aprenden lo que se espera de ellos al finalizar cada nivel; cuál es el grado de equidad o inequidad en el logro de esos aprendizajes; cómo evolucionan a lo largo de los años los niveles de logro y la equidad en el acceso al conocimiento de los distintos grupos sociales; cuánto y cómo las desigualdades sociales inciden sobre las oportunidades de aprendizaje de los estudiantes; cómo influyen las prácticas escolares y las condiciones de la enseñanza (situación de los maestros, recursos, tiempo disponible, etc.) en los aprendizajes estudiantiles; qué efectos tienen en los logros educativos las inversiones, los cambios en el sistema, los cambios curriculares, la compra de material educativo, etc.
- *Aportan información importante a diversos actores sociales.* En la medida que el sistema produzca y comunique la información arriba citada puede ser un instrumento clave de mejora, que enriquezca la comprensión de la situación educativa y la toma de decisiones de *autoridades y encargados de formular políticas educativas, los directores de escuelas y docentes, los supervisores y encargados de la formación de docentes, los padres y madres de los estudiantes, y la ciudadanía* –dándole a cada grupo de actores herramientas intelectuales valiosas para tomar decisiones y participar activamente en la comunidad educativa según el rol que les corresponde.
- *Conducen a discusiones informadas sobre qué aspectos del currículo son exigibles a todos los estudiantes y a definir con claridad qué es lo que todos deberían haber aprendido al final de cada ciclo educativo.* La mayoría de los currículos latinoamericanos son largas listas de objetivos y contenidos, todos deseables, pero no realizables en la práctica. Las macro-evaluaciones obligan a establecer qué es o no fundamental y, por tanto, lo que los estudiantes deben saber y ser capaces de hacer. Estas definiciones se pueden designar con distintos términos: competencias fundamentales, metas de aprendizaje, indicadores de logro, niveles de desempeño o estándares, entre otros.

Después de revisar los significativos aportes de la macro-evaluación, conviene también indicar sus limitaciones para evitar distorsiones o simplificaciones en la interpretación. Citando nuevamente los agudos análisis de Ravela et al. (2008, p.48), las limitaciones que no hay que perder de vista son:

- *Una macro-evaluación estandarizada aporta información fundamental e indispensable sobre la calidad educativa, pero no es un indicador completo de la misma.* Hay saberes, actitudes, valores y aprendizajes que no pueden (por la dificultad de medirlos de esta manera) o no deben (porque son propios del contexto local y no exigibles nacionalmente) formar parte de ella.
- *La macro-evaluación estandarizada de los aprendizajes es un componente esencial de un sistema integral de evaluación educativa, pero no es el único tipo de evaluación relevante.* También tienen importancia la evaluación que los docentes realizan en el aula, la evaluación del desempeño docente, la evaluación de los centros educativos, la de las políticas educativas, la del uso de los recursos, la del currículo, etc.
- *La macro-evaluación estandarizada externa sólo tendrá efectos positivos sobre la educación si es concebida, percibida y empleada como un mecanismo de responsabilidad pública de todos los actores vinculados al quehacer educativo.* Existe el riesgo de que se implementen evaluaciones, pero después no se tomen acciones para resolver los problemas que ellas detectan. A veces las autoridades informan de los resultados y dejan la responsabilidad de los problemas identificados a escuelas y familias, como si fuera un asunto privado; otras veces se responsabiliza a los docentes, quienes a su vez transfieren la responsabilidad a los padres o al contexto. Hay que evitar entrar en una dinámica de culpabilización mutua y, por el contrario, construir una lógica de responsabilidad educativa compartida.
- *La evaluación en general es condición necesaria pero no suficiente para mejorar la educación.* La evaluación es solo uno de los varios elementos claves de la política educativa que deben estar alineados. Otros son la formación docente inicial y en servicio, la carrera y condiciones del trabajo docente, la gestión escolar, la supervisión, los currículos, los materiales educativos, la inversión de recursos proporcionales a las necesidades de las distintas poblaciones y la acción de los responsables educativos para resolver los problemas encontrados, entre otros.

Vistos los valiosos aportes que hace la macro-evaluación de los aprendizajes, el tema de la presente investigación es muy importante para la educación en cualquier lugar del mundo. Pero... ¿por qué necesita especial atención de la educación costarricense actual? Porque a la fecha de conclusión de la presente investigación (febrero de 2023), el Consejo Superior de Educación, después de un año de incertidumbre, aprobó el nuevo abordaje académico de la macro-evaluación propuesto por la actual administración. Se anunció que el mismo será puesto en práctica a partir de marzo 2023, aunque después el comienzo de las pruebas se retrasó a abril 2023. La situación del país en este sentido sigue siendo débil y requiere mayores precisiones técnicas.

Poco tiempo atrás, al examinar la política educativa de Costa Rica, la Organización para la Cooperación y el Desarrollo Económico, OECD, expresaba que *“en los últimos años el país ha hecho importantes progresos en cuanto a desarrollar un sistema nacional de evaluación”*. Destacaba hitos significativos como *“la introducción de un marco para monitorear la calidad de los procesos educativos (Modelo de Evaluación de la Calidad de la Educación Costarricense, MECEC); el establecimiento de una unidad dentro del Ministerio dedicada a la administración y evaluación de la calidad, y la introducción de una evaluación nacional en gran escala del aprendizaje estudiantil al final de la educación primaria y la educación secundaria inferior. Como resultado de estas iniciativas, ahora Costa Rica tiene mejor información acerca de cómo funciona el sistema educativo y comienza a poner más atención en la calidad del proceso y los resultados educativos”* (OECD, 2017, ps. 126-27).

Lamentablemente, lo que entonces la OECD valoraba como progresos importantes e hitos significativos de un sistema nacional de evaluación han desaparecido, o están seriamente cuestionados en su calidad y eficacia. Por una parte, la evaluación nacional de gran escala que entonces empezaba a perfilarse –las pruebas *FARO, Fortalecimiento de Aprendizajes para la Renovación de Oportunidades*– fueron derogadas en 2022, a poco de aplicarse por primera vez. Por otra parte, el Modelo de Evaluación de la Calidad de la Educación Costarricense, MECEC, planteado por el Ministerio de Educación Pública, MEP, así como la unidad dentro del Ministerio dedicada a la administración y evaluación de la calidad, fueron seriamente cuestionadas por una cantidad de falencias en un Informe de Auditoría de la Contraloría General de la República de agosto de 2022. En ese Informe se detectaron como grandes problemas la ausencia de una visión estratégica de la evaluación de la calidad de la educación; debilidades en la estructura y articulación del Sistema Nacional de Evaluación de la Calidad de la Educación, SNECE, que limitaban la implementación de sus funciones, y ausencia de información integrada sobre la calidad de la educación. (Contraloría General de la República, 2022, ps.10-24).

En consecuencia, el sistema educativo costarricense enfrenta distintas limitaciones en cuanto a la macro-evaluación de los aprendizajes. Es, pues, un momento oportuno para analizar experiencias valiosas de macro-evaluación que se llevan adelante en el país y fuera de él, que puedan contribuir a inspirar y facilitar decisiones que Costa Rica tiene por delante.

Por lo anterior, esta investigación se propone dos objetivos. Primero, sistematizar las principales características y lecciones aprendidas de las buenas experiencias en materia de macro-evaluación que existen en Costa Rica y en otros países con modelos educativos exitosos y con avances en este campo en América Latina y la OCDE. Segundo, elaborar recomendaciones y avanzar en una propuesta de hoja de ruta para que el país mejore en los próximos años la gobernanza de los procesos de macro-evaluación y el uso de la información para incidir en el trabajo de aula de los docentes y en la mejora de los aprendizajes estudiantiles.

Algunas precisiones sobre evaluación educativa

Conviene comenzar clarificando algunos términos de uso frecuente en la materia para evitar confusiones o malos entendidos posteriores.

Según sus objetivos, la literatura especializada diferencia entre *evaluación sumativa* y *formativa* y muchos autores también distinguen entre *evaluación formativa* y *diagnóstica*. La *evaluación sumativa* o *evaluación de lo aprendido* (generalmente referida a estudiantes, aunque también puede referirse a docentes, si se los está evaluando) busca sintetizar el aprendizaje que una persona ha logrado a fin de llevar un registro, calificarlo o certificar sus logros. (EPPI, 2002) Cuando se trata de una evaluación que conduce a certificar cierto nivel de aprendizajes mediante un documento oficial extendido por una autoridad educativa se denomina *evaluación de certificación o acreditación*.

La evaluación formativa o *evaluación para aprender* busca examinar aspectos de un aprendizaje en marcha –o sea, que una persona está haciendo en esos momentos– a fin de profundizarlo y orientar los pasos que siguen. La *evaluación diagnóstica* es un *tipo de evaluación formativa* que suele realizarse al inicio de un nuevo proceso de aprendizaje para una persona con el fin de identificar un punto de partida o línea de base para guiar ese aprendizaje y desarrollar un programa adecuado de enseñanza o, en términos utilizados actualmente, de mediación pedagógica. La evaluación diagnóstica también puede servir para identificar estudiantes que están en riesgo de fracaso escolar a efectos de descubrir las fuentes de sus dificultades de aprendizaje y diseñar una intervención pedagógica complementaria o remedial. (OECD, 2017, p. 140-41)

Según su alcance, la literatura distingue entre *macro-evaluación* y *micro-evaluación*. La *macro-evaluación*, como se explicó en la sección anterior, se aplica a nivel del país y puede tener distintos objetivos a fin de satisfacer distintas necesidades identificadas por el sistema educativo: puede ser sumativa, de certificación o acreditación, formativa, o diagnóstica. Utiliza como instrumento de medición pruebas estandarizadas construidas por especialistas y validadas, que aseguran que las puntuaciones de conjuntos grandes de individuos se pueden comparar entre sí y las inferencias que se pueden extraer de ellas son válidas.

Siempre dentro de *la macro-evaluación* y según la población a la que se aplican, las pruebas estandarizadas se clasifican en *pruebas censales* o *muestrales*. *Las pruebas censales* se aplican a toda la población estudiantil y las escuelas de un país o territorio (de allí su nombre, porque son similares a un censo), mientras que las *pruebas muestrales* se aplican a una muestra escogida de la población estudiantil y de escuelas.

La macro-evaluación es un componente central del sistema de evaluación de un país, pero siempre se complementa con la micro-evaluación. *La micro-evaluación* es la que aplican los docentes a grupos pequeños en contextos reducidos, generalmente a nivel de su propia aula de clase. También puede tener distintos objetivos según las necesidades del docente o del centro escolar –puede ser sumativa, formativa o diagnóstica. Sus instrumentos son pruebas construidas para cada el aula, por lo común por el respectivo docente, y las puntuaciones se

adaptan exclusivamente al grupo y a la situación específica del aula o la escuela. El Cuadro 1 resume las principales diferencias entre los tipos de instrumentos que se utilizan en la macro y en la micro-evaluación.

Cuadro 1

Diferencias entre pruebas estandarizadas y pruebas construidas en el aula

Aspecto	Pruebas estandarizadas	Pruebas en el aula
Resultados de aprendizaje y contenidos medidos	Los resultados y contenidos objeto de medición deben ser comunes para todos los centros educativos. Las pruebas rara vez enfatizan aspectos locales del currículo.	Se consideran aspectos específicos del aprendizaje generado a partir de un determinado proceso de mediación pedagógica y una dinámica particular en el aula.
Calidad de los ítems	En general, si se siguen los procedimientos establecidos por la Psicometría y la medición educativa, la calidad técnica es alta. Los ítems son elaborados por especialistas, sometidos a pruebas piloto y seleccionados con base en su efectividad para lograr una medición adecuada.	La calidad de los ítems es desconocida, o al menos no suele estar documentada. Típicamente la calidad es más baja que la de las pruebas estandarizadas, por el poco tiempo del que dispone el docente y la falta de conocimientos especializados.
Confiabilidad	Alta confiabilidad, comúnmente mayor a 0,8 (tanto en la TCT –teoría clásica de los tests– como en el modelo de Rasch).	Por lo general desconocida; puede ser alta si la prueba es cuidadosamente construida. Rara vez se calculan medidas empíricas.
Administración y calificación	Procedimientos estandarizados (uniformes). Se brindan instrucciones específicas. Existen protocolos de administración y calificación.	Se favorecen los procedimientos uniformes, pero hay laxitud. Muchas veces los protocolos no existen o son poco específicos.
Interpretación de puntajes	Los puntajes se asocian a niveles de desempeño. Existen manuales y otras guías que ayudan a usar las pruebas e interpretar sus resultados.	Las comparaciones de los puntajes y las interpretaciones se limitan a la situación específica del centro educativo o el aula.
Número de examinados	Aplicación masiva, usualmente varios cientos o miles.	Aplicación reducida, grupos pequeños en un contexto restringido.

Fuente: Montero, Rojas, Zamora y Rodino Pierri, 2012, con base en Gronlund y Linn, 1990, en Estado de la Educación IV, 2013, p.259.

Buenas prácticas nacionales e internacionales sobre factores centrales de la macro-evaluación

La macro-evaluación educativa es un proceso académico y técnicamente complejo, en el que entran en juego una cantidad de factores o variables. Entre los principales están, por ejemplo, sus posibles objetos de estudio (estudiantes, docentes, personal directivo o centros escolares); el actor institucional que la diseña y lleva adelante (ministerios y secretarías de educación, o entidades independientes); sus objetivos; las técnicas e instrumentos de medición que elige emplear, y la o las maneras en que utilizan sus resultados para contribuir a mejorar su objeto de estudio —por ej. el desempeño estudiantil, docente o del centro escolar dentro del sistema educativo—. De aquí que identificar y seleccionar buenas prácticas en la materia difícilmente incluirá propuestas únicas, que cubran todos los factores importantes al mismo tiempo y que a la vez sean pertinentes para las principales necesidades y aspiraciones educativas de Costa Rica.

Las buenas prácticas —también llamadas prácticas exitosas— nacionales e internacionales que se recogen en este documento se refieren exclusivamente a la macro-evaluación de los desempeños estudiantiles, e ilustran cómo se pueden manejar eficazmente los factores centrales del proceso: (1) su forma de gobernanza y organización institucional; (2) sus diferentes objetivos; (3) las técnicas e instrumentos empleados para medir los aprendizajes estudiantiles, y (4) el uso que se hace de los resultados obtenidos para mejorar la calidad de los aprendizajes. Por lo dicho, al identificar buenas prácticas sobre diferentes factores del proceso evaluativo, hay casos que se repiten —sean países, unidades territoriales dentro de países u organismos que conducen procesos de evaluación internacionales.

Gobernanza y organización institucional

La noción de *gobernanza y organización de la macro-evaluación* se enfoca en la entidad que conduce y ejecuta el proceso en una unidad territorial determinada. Esto supone examinar en especial cuál es esa entidad y que relación institucional tiene con el respectivo ministerio o secretaría de educación pública del país (si son entidades independientes, o bien una es parte de la otra y con qué grado de dependencia) y cuáles son las tareas y responsabilidades de cada una respecto al diseño y ejecución de la evaluación.

En este sentido, la literatura especializada coincide en que la mejor práctica es que un país —o una unidad territorial menor como una provincia o estado, si tiene a su cargo la educación— ***Cree una entidad o agencia especializada con la función específica de diseñar y conducir el proceso de macro-evaluación, que sea autónoma de las autoridades educativas*** (del ministerio o secretaría de educación nacional, provincial o estadual según corresponda). (OECD, 2013 y 2017)

Esta práctica representa un enfoque novedoso de gobernanza de la macro-evaluación que trae ventajas muy importantes para el proceso técnico de evaluación en sí mismo y para los sistemas educativos que la emplean. Comenzó a implementarse en el mundo entre la última década del Siglo XX y la primera del Siglo XXI, al principio con avances lentos, para irse afirmando y extendiéndose en la última década. Esto ocurre en contextos donde las *autoridades educativas*

conservan el liderazgo de fijar la estrategia y la política educativa y mantienen un rol en implementar las conclusiones de la evaluación, mientras las *agencias de evaluación* brindan un juicio independiente sobre el logro de los objetivos educativos. Establecer estas agencias autónomas implica darles independencia política y financiera respecto de las autoridades educativas y una presencia importante de especialistas en sus cuerpos directivos. El objetivo debe ser afianzar a estas agencias como voces autorizadas en sus áreas de competencia, muy creíbles por su capacidad y experticia técnica, y haciendo recomendaciones para llevar a la práctica los resultados de la evaluación en el país. (OECD, 2013, p.120)

Las ventajas de esta práctica son varias. Para empezar, le otorga a la macro-evaluación cierto grado de independencia frente a las autoridades educativas (que, de otra manera, actúan en este sentido como “juez y parte”) y fortalece la rendición de cuentas pública de esas autoridades respecto a los resultados del sistema educativo. En segundo lugar, permite contar con una concentración de experticia especializada en la materia, indispensable para desarrollarla de manera cada vez más adecuada y precisa. En tercer lugar, apoyada en la ventaja anterior, abre la posibilidad de construir una sólida capacidad institucional para aplicar políticas de macro-evaluación. (OECD, 2013, p.39)

A las ventajas apuntadas por la OECD esta investigadora agrega, tomando en cuenta experiencias latinoamericanas recientes, que independizar la evaluación educativa de los ministerios de educación puede evitar o reducir vaivenes abruptos en la materia por razones políticas o ideológicas en lugar de razones técnicas, derivadas de cambios en la administración gubernamental, tal como ocurrió en Costa Rica con la suspensión de las pruebas FARO, *Fortalecimiento de Aprendizajes para la Renovación de Oportunidades* en 2022.

Es pertinente destacar que la OECD, en su amplio informe de hace pocos años sobre la educación en Costa Rica, recomienda al país precisamente la creación de una agencia independiente de evaluación (OECD, 2017, p. 131). El informe recuerda que la idea no es inédita ya que Costa Rica creó un órgano de este tipo –el *Centro Nacional de Evaluación de la Educación, CENE-EDU*–, desconcentrado del Consejo Superior de Educación, en 1997 (Decreto Ejecutivo 26014 del 12/5/97)–, el cual desafortunadamente fue eliminado apenas un año después (Decreto Ejecutivo 27089 del 28/5/98). (Chavarría Navas, 2010, p. 73-4)

¿Qué países o territorios adoptaron esta buena práctica? Actualmente su número en el mundo es muy nutrido. Citaremos los casos más destacados en nuestro continente, describiéndolos con algún detalle, y después haremos una mención puntual de los países de la OECD que también la han adoptado en el Cuadro 2.

Chile. Es el caso latinoamericano más reconocido en cuanto a progresos en evaluación. Si bien el país mide la calidad educativa desde hace 30 años, a partir de la Ley 20529 de 2011 cuenta con un *Sistema de Aseguramiento de la Calidad de la educación parvularia, básica y media, SAC*, cuya misión es lograr la mayor equidad y calidad en la educación escolar. El SAC está formado por cuatro instituciones con funciones específicas: (a) el *Ministerio de Educación*, órgano rector a cargo de diseñar e implementar las políticas educacionales para todo el sistema educativo; (b) la *Agencia de Calidad de la Educación, ACE*, o simplemente la *Agencia*, responsable de evaluar los logros de aprendizaje, los indicadores de desarrollo personal y social y el desempeño de las

escuelas para orientarlas en su quehacer institucional y pedagógico e informar a la comunidad escolar de estos procesos; (c) el *Consejo Nacional de Educación, CNE*, que aprueba y comunica los instrumentos curriculares y de evaluación para la educación escolar que le presentan el Ministerio y la Agencia, y (d) la *Superintendencia de Educación*, cuya tarea es fiscalizar el uso de recursos y el cumplimiento de la normativa educativa, atender denuncias y reclamos y establecer sanciones, de manera que las escuelas cumplan con las necesidades estudiantiles. (Agencia de Calidad de la Educación, 2018; Lagos, 2021)

La Agencia es colegiada igual que el CNE, y ambas tienen una gobernanza que les permite tomar distancia del calendario electoral. Sus cargos, igual que los del CNE y de la Superintendencia de Educación, son concursables, no políticos, con una duración que no coincide con los ciclos electorales. La Agencia se financia con parte del presupuesto nacional, pero no del presupuesto del Ministerio de Educación. Todas estas previsiones se adoptaron al crearse el Sistema de Aseguramiento de la Calidad, SAC, precisamente para preservar la independencia de la Agencia respecto al Ministerio de Educación y sus autoridades esencialmente políticas. (E: Severin, 2022)

Las funciones de la Agencia son evaluar, orientar e informar por lo que, aparte de brindar resultados de aprendizaje, informa a las escuelas sobre los factores asociados a esos resultados. Aquellas que presentan los desempeños más bajos reciben visitas integrales de evaluación y orientación que se basan en los Estándares Indicativos de Desempeño elaborados por el Ministerio de Educación –que abordan estándares de liderazgo, gestión pedagógica, formación y convivencia y gestión de recursos. Junto a ampliar el foco de la evaluación incluyendo los procesos de la escuela, la Agencia amplió la mirada sobre los aprendizajes para que estos no se restrinjan a los aspectos académicos sino consideren también el desarrollo personal y social de los estudiantes. Actualmente, con base en los acuerdos y lineamientos estipulados en el Plan de Evaluaciones Nacionales e Internacionales 2016-2020, trabaja en el desarrollo de un nuevo Sistema Nacional de Evaluación de Aprendizajes con nuevos componentes, más amplio y equilibrado, y entrega información que responde a las necesidades de distintos actores educativos. (Agencia de Calidad de la Educación, 2018, p.9-11; <https://www.agenciaeducacion.cl>)

La Agencia utiliza la prueba SIMCE –Sistema de Medición de la Calidad de la Educación– que empezó en el país desde fines del siglo pasado. Es una prueba censal sumativa, de contenido curricular, que se aplica en 4º, 6º, 8º y 10º grados de primaria y secundaria, básicamente en las asignaturas Lenguaje, Matemáticas, Ciencias e Historia. (E: Severin, 2022) Cuando se dan a conocer resultados individuales, se utiliza una escala de logro dividida en cuatro *niveles de desempeño* (bajo, medio bajo, medio alto y bueno) que detallan lo que el estudiante es capaz de hacer en la materia evaluada. (E: Cortázar, 2022) Esta práctica, que Ferrer y Fiszbein describen como “*una metodología más moderna, sofisticada y útil para futuras intervenciones pedagógicas de mejoramiento*” (2015, p.10), se impone en la actualidad en muchos países por encima del uso de calificaciones puntuales.

Uruguay. La entidad a cargo de la macro-evaluación es el *Instituto Nacional de Evaluación Educativa, INEED*, creado por la Ley General de Educación 18.437 de 2009, que empezó a funcionar en 2012. Fue constituido como una persona pública de derecho privado (no estatal),

lo que significa que es plenamente autónoma en su accionar. Su autoridad es una Comisión Directiva de tres miembros que son elegidas por el gobierno, mientras que sus funcionarios son privados. Se vincula con el Poder Ejecutivo a través del Ministerio de Educación y Cultura, pero tiene presupuesto propio establecido por ley. (www.ineed.edu.uy/institucional, consultado el 26/9/22; E: Lasida, 2022)

Su misión es evaluar la calidad de la educación nacional a través de estudios específicos (pruebas estandarizadas) y el desarrollo de líneas de investigación educativa. Para ello debe cumplir diversas funciones: centrarse en los niveles inicial, primario y medio; aportar información que contribuya a garantizar el derecho a la educación de los educandos; dar a conocer el cumplimiento de los objetivos y metas establecidos por los diferentes organismos e instituciones educativas del país; producir conocimiento sobre los procesos de evaluación; aportar información acerca de los aprendizajes de los educandos; proponer criterios y modalidades en los procesos evaluativos del Sistema Nacional de Educación en los niveles inicial, primario y medio (obsérvese que esta función se refiere a un apoyo a la evaluación formativa no estandarizada que siguen aplicando las escuelas, como un sistema paralelo al del INEED, E: Lasida, 2022); asesorar al Sistema Nacional de Educación en cuanto a la participación en instancias internacionales de evaluación, y contribuir con las entidades estatales o privadas vinculadas a la educación prestando servicios de evaluación. (www.ineed.edu.uy/institucional, consultado el 26/9/22). Respecto a la última misión encomendada por la ley, nótese que se establece que la educación nacional incluye tanto la pública como la privada. (E: Lasida, 2022)

La prueba de macro-evaluación uruguaya se llama ARISTAS y se aplica para 3º y 6º grados de primaria y para 3º de media básica, cada 3 años. Son estandarizadas y muestrales. Evalúan el currículo en materia de Lengua, Matemáticas y Habilidades Socio-Emocionales. Si bien el currículo nacional es muy antiguo y muy contenidista, por lo cual está en proceso de cambio, el INEED realizó el proceso de convertir esos contenidos en *aprendizajes* (formulación igual o muy próximo a *competencias*), con lo cual le introdujo una valiosa actualización. Los resultados de ARISTAS diferencian cinco niveles de logro o *niveles de desempeño*. (E: Lasida, 2022)

Además de ser una entidad especializada e independiente de las autoridades nacionales de educación, el INEED es muy reconocido por su exitoso desarrollo de instrumentos de evaluación y por la aplicación de los resultados al trabajo de aula con respaldo en la tecnología informática. Veremos estas otras buenas prácticas en secciones siguientes.

Colombia. El *Instituto Colombiano para la Evaluación de la Educación, ICFES*, es la entidad a cargo de las actividades de evaluación en todos los niveles educativos, incluyendo los Exámenes de Estado al final del nivel secundario para el ingreso a la universidad. Fue creado en 1968, pero tuvo una importante reforma estructural en 2009 que lo convirtió en un organismo autónomo adscrito al Ministerio de Educación Nacional. A partir de entonces se creó también un nuevo sistema de exámenes, las pruebas SABER, que abarcan todo el espectro de la educación en el país, desde la básica, pasando por la media y la técnica y tecnológica, y llegan hasta la educación superior.

Las principales funciones del ICFES son: establecer las metodologías y procedimientos que guían la evaluación externa de la calidad de la educación; desarrollar la fundamentación teórica, diseñar, elaborar y aplicar instrumentos de evaluación de la calidad de la educación, dirigidos a los estudiantes de los niveles de educación básica, media y superior, de acuerdo con las orientaciones del Ministerio de Educación Nacional; diseñar, implementar, administrar y mantener actualizadas las bases de datos con la información de los resultados de las pruebas aplicadas y los factores asociados; organizar y administrar el banco de pruebas y preguntas, según niveles educativos y programas; diseñar e implementar el procesamiento de información y la divulgación de resultados de las evaluaciones realizadas; realizar estudios e investigaciones en el campo de la evaluación de la calidad de la educación; impulsar la cultura de la evaluación mediante la difusión de los resultados educativos y el análisis acerca de los factores que inciden en los mismos, y el desarrollo de actividades de formación dentro de su competencia, y apoyar al Ministerio de Educación en realizar los Exámenes de Estado.

Las pruebas de evaluación SABER están así organizadas:

- SABER 3º, 5º y 9º monitorean el desarrollo de competencias básicas en los estudiantes de la educación básica mediante cuatro pruebas –Lenguaje, Ciencias Sociales, Ciencias Naturales y Matemáticas, más un cuestionario de contexto.
- SABER 11º evalúa las competencias de los estudiantes que están por terminar el grado undécimo de la educación media y consta de cinco pruebas –Lectura Crítica, Matemáticas, Ciencias Sociales y Ciudadanas, Ciencias Naturales e Inglés, más un cuestionario socioeconómico. Esta prueba es requisito para acceder a la educación superior y es utilizada como criterio de admisión por varias universidades. (corresponde a los anteriores Exámenes de Estado, nombre que todavía se utiliza).
- SABER T y T evalúan las competencias genéricas y específicas de los estudiantes de programas técnicos y tecnólogos que aprobaron el 75% de los créditos académicos por medio de cinco módulos de preguntas centradas en competencias y un cuestionario socioeconómico. Incluye tres módulos sobre contenidos específicos que los estudiantes presentan según su área de formación profesional (Ensamblaje, mantenimiento y operación de maquinaria y equipos; Mantenimiento e instalación de hardware y software, y Promoción de la salud y prevención de la enfermedad).
- SABER PRO son un examen oficial y obligatorio que comprueba las competencias de los estudiantes que cursan el último año de los programas de pregrado en las instituciones de educación superior. Se realiza en dos sesiones, una obligatoria de competencias genéricas (Lectura Crítica, Razonamiento Cuantitativo, Competencias Ciudadanas, Comunicación Escrita e Inglés) y otra para estudiantes inscritos por su institución para presentar módulos específicos de su área de formación.

Nótese que en Colombia no hay un currículo, sino estándares e indicadores nacionales como instrumentos básicos del aprendizaje. Las pruebas nacionales dan los resultados en una escala de 4 *niveles de desempeño*, de manera similar a como lo hacen Chile y Uruguay. (www.icfes.gov.co y <https://es.wikipedia.org/wiki/ICFES>, consultados el 14/10/22; E: Figueroa, 2022)

Perú. Tiene una práctica intermedia pero exitosa respecto a la gobernanza de la evaluación. En este caso, como en Costa Rica, las evaluaciones nacionales de logros de aprendizaje en instituciones públicas y privadas son implementadas por el Ministerio de Educación a través de la *Oficina de Medición de la Calidad, OMC*. Pero, si bien se trata de una instancia interna del Ministerio, goza de autonomía, además de gran legitimidad y prestigio. (E: Cuenca, 2022)

El país tiene experiencia con pruebas nacionales desde 1990 (si bien se suspendieron entre 1998 y 2003). Eran *pruebas muestrales*, que en 2007 pasaron a ser *censales*. Desde entonces existe un debate sobre su utilidad y si deberían regresar a ser muestrales a fin de devolver información al sistema en vez de a las aulas y sobre cada estudiante, porque esto último produce dos efectos perversos: crear bonos e incentivos para los docentes, quienes se centran en “enseñar para la prueba”, y generar competición entre distintas regiones del país. (E: Cuenca, 2022)

Más allá de la discusión sobre cuál prueba preferir, la evaluación en el país mantiene legitimidad y genera confianza en sus resultados. La relativa independencia y la experticia técnica de la OMC permitieron que en los últimos años se dieran dos fenómenos valiosos: a partir de las evaluaciones se crearon y sostuvieron *políticas públicas* pensadas de manera más orgánica y se construyó una *narrativa positiva sobre la evaluación*. Ambos ayudaron a mejorar la evaluación, a generar materiales y a marcar una ruta de aprendizaje continuo. (E: Cuenca, 2022)

En síntesis, las funciones y actividades de distintas agencias de evaluación latinoamericanas independientes ilustran –aunque en grado variable–, las ventajas de contar este tipo de entidad para impulsar de manera constante la evaluación educativa. Ello se concreta en crecientes desarrollos técnicos, capacidades institucionales, resultados robustos en materia de evaluación e impactos positivos en los aprendizajes estudiantiles.¹

Cuadro 2

Agencias intermedias especializadas con un rol en la macro-evaluación por país de la OECD (2012)

País	Agencia especializada
Australia (a nivel federal)	Autoridad Australiana de Currículo, Evaluación e Información (ACARA) (creada en 2009) Instituto Australiano de Enseñanza y Liderazgo Escolar (AITSL) (creado en 2010)
Austria (a nivel federal)	Instituto Federal para la Investigación Educativa, la Innovación y el Desarrollo del Sistema Escolar Austríaco (BIFIE) (creado en 2008)
Bélgica (Comunidad Flamenca)	Agencia para el Aseguramiento de la Calidad de la Educación y la Capacitación (AKOV) (creada en 2009) Inspección Flamenca de la Educación (creado en 1991)
Bélgica	Servicio General de Inspección (asociada con el Ministro de la Comunidad Francesa) Comisión de Monitoreo del Sistema de Educación (creada en 2002)

¹ Cabe aclarar que las agencias autónomas de evaluación en América Latina y la mayoría en otros países de la OECD también incluyen entre sus funciones la conducción de algún tipo de evaluación o auto-evaluación de docentes y establecimientos educativos, con variaciones entre sí. No se describen en este Informe porque exceden los objetivos de la investigación.

País	Agencia especializada
(Comunidad Francesa)	
Canadá (a nivel federal)	Consejo de Ministros de Gobierno, Canadá (Programa Pan-Canadiense de Evaluación)
Chile	Agencia de Calidad de la Educación (creada en 2011, comenzó a funcionar en 2012)
República Checa	Inspección de Escuelas Checas (creado en 1991, reformado en 2005) Centro para la Evaluación de Logros Educativos (CERMAT) (dependiente del Ministerio de Educación, Juventud y Deportes) (creado 1999)
Dinamarca	Agencia de Calidad y Supervisión (creada en 2011) Consejo para la Evaluación y el Desarrollo de Calidad de la Educación Primaria y Secundaria Inferior (creado en 2006) Instituto Danés de Evaluación (creado en 1999)
Finlandia	Consejo Nacional de Educación (creado en 1991) Consejo de Evaluación de la Educación
Francia	Inspección General de Educación Nacional (parte del Ministerio de Educación) Inspección General de la Administración de la Educación Nacional y la Investigación (parte del Ministerio de Educación)
Hungría	La Autoridad Educacional (creada en 2006)
Islandia	Instituto de Pruebas Educativas
Irlanda	Inspección del Departamento de Educación y Habilidades Consejo Nacional de Currículo y Evaluación (NCCA) (creado en 2001) Comisión de Exámenes de Estado (entidad separada dependiente del Departamento de Educación y Habilidades) (creada en 2003) Centro de Investigación Educativa (creado en 1966)
Italia	Instituto Nacional para la Evaluación del Sistema Educativo (INVALSI) (creado en 2004) Instituto Nacional de Investigación, Experimentación y Desarrollo Educacional (INDIRE)
Corea	Instituto Coreano de Currículo y Evaluación (KICE) (creado en 1998) Instituto Coreano de Desarrollo Educativo (KEDI) (creado en 1972)
Luxemburgo	Agencia para el Desarrollo de la Calidad en las Escuelas (integrada al Ministerio Nacional de Educación y Capacitación Vocacional) (creado en 2009)
México	Comisión Nacional para la Mejora Continua (creada en 2019, descentralizada pero no autónoma)(*) Instituto Nacional de Evaluación Educativa (INEE) (creado en 2002, transformado en 2018) (*)
Holanda	Inspección Holandesa de Educación Consejo de Reconocimientos (CVE)

País	Agencia especializada
Nueva Zelandia	Oficina de Evaluación Educativa Autoridad de Certificaciones de Nueva Zelandia (creada en 1989)
Noruega	Dirección de Educación y Capacitación (agencia ejecutiva del Ministerio de Educación e Investigación) (creado en 2004)
Polonia	Comisión Central Examinadora
Portugal	Inspección General de Educación y Ciencia (integrada en la estructura del Ministerio de Educación y Ciencia (creada en 1979)
República Eslovaca	Instituto Nacional para la Medición Educativa Certificada (NÚCEM) (creada en 2008) Inspección de Escuelas Estatales
Eslovenia	Consejo para la Calidad y la Evaluación (Cuerpo Asesor del Ministerio de Educación y Deporte, creado en 1999) Centro Nacional Examinador (creado en 1993) Inspección de la República de Eslovenia para la Educación y el Deporte (afiliado al Ministerio de Educación y Deporte) Instituto Nacional de Educación (creado en 1956)
Suecia	Inspección Sueca de Escuelas (creada en 2008) Agencia Nacional de Educación (creada en 1991)
Reino Unido (Irlanda del Norte)	Consejo del Currículo, Examinaciones y Evaluación (creado en 1993) Inspección de Educación y Capacitación (parte del Departamento de Educación)

Fuente: OECD, 2013, p. 69, basado informes proporcionados por los países. Traducido por A. M. Rodino.

(*) Datos actualizados por A. M. Rodino a partir de la entrevista con F. Miranda López y D. Monroy Magaldi, México, 2022.

Antes de concluir la sección es necesario aclarar por qué el cuadro anterior contiene un agregado de datos actualizados para **México**, incorporados por la investigadora, que no se encuentran en la fuente original de la OECD, 2013. La aclaración explica a la vez la razón por la cual este país no está incluido entre los ejemplos de buenas prácticas latinoamericanas citadas en este Informe.

México contaba desde 2002 con el *Instituto Nacional de Evaluación Educativa, INEE* como órgano constitucional autónomo a cargo de la evaluación nacional del desempeño estudiantil, de docentes y de centros educativos. El INEE tenía un gran desarrollo y experiencia en la materia, al punto que la publicación especializada de la OECD, 2013, lo referencia a menudo como una práctica exitosa en varios sentidos. Pero en 2018, como resultado de numerosas manifestaciones docentes de protesta, hubo un fuerte debate parlamentario y el INEE desapareció. En su lugar se creó la *Comisión Nacional para la Mejora Continua de la Educación* como cuerpo descentralizado, pero no autónomo, y la evaluación de nivel nacional pasó a definirse como diagnóstica, voluntaria, participativa y comunitaria. Se redujeron las pruebas estandarizadas a un solo instrumento muestral para evaluar aprendizajes fundamentales en Lectura, Matemática y Formación Cívica y Ética. El cambio fue importante desde el punto de vista político pues las funciones originales se diversificaron al agregársele otras de innovación

educativa, formación de docentes y estudios especializados. (E: Miranda López y Monroy Magaldi, 2022). Dado que en la actualidad la *Comisión* está terminando de desarrollar y pilotear un nuevo modelo de evaluación, que no se ha piloteado, es prematuro identificar prácticas exitosas en su proceso, motivo por el cual el país no se referencia en el presente documento.

Objetivos

Los objetivos de la evaluación son los que les dan nombre a sus diferentes variantes. Según lo expuesto en la sección 2, pueden ser (a) *objetivos formativos* para retroalimentar los aprendizajes con el fin de mejorarlos –y dentro de ellos, como un subtipo, *diagnósticos* o (b) *objetivos sumativos*, orientados a calificar los aprendizajes, dentro de los cuales se incluyen los *objetivos de certificación o acreditación* de aprendizajes adquiridos en el ciclo o nivel que se evalúa.

Los objetivos que se fijen para la evaluación en un país son decisiones de política educativa, es decir, son definidos por las autoridades rectoras del sistema de educación pública nacional por distintas razones, que pueden ser desde técnico-educativas hasta político-ideológicas para satisfacer a determinados grupos de presión (por ej. el propio partido gobernante, sindicatos de educadores o representantes de académicos, empresarios o medios masivos).

El criterio que predomina en instituciones especializadas como la OECD es que *“la evaluación en educación es una cuestión de encontrar equilibrios entre sus objetivos formativos y sumativos”* (E: Santiago, 2022). Ella recomienda, pues, tener un sistema que pueda cumplir todas las funciones de la evaluación, pues un único modelo no permite alcanzar sus distintos objetivos, pertinentes para distintas situaciones, así como tampoco puede hacerlo un solo instrumento de evaluación. Las políticas públicas en la materia tienen que considerar el contexto de cada país, pues algunos, como los europeos, tienen una larga historia y cultura de evaluación mientras que otros, como los latinoamericanos, tienden a percibirla como una forma de control de normas. Por ello es recomendable tener simultáneamente dos tipos de evaluación con sus correspondientes instrumentos: una para identificar mejoras necesarias en la enseñanza –que por lo común es más interna de las escuelas– y otra para calificar y rendir cuenta de los aprendizajes alcanzados –externa a las escuelas, de alcance nacional (E: Santiago, 2022).

Sin perjuicio de lo anterior, ambas evaluaciones tienen que relacionarse: la macro-evaluación externa a las escuelas debe conectarse con el desarrollo de las capacidades locales de los centros educativos. A diferencia de Europa, esta tradición es escasa y débil en América Latina. Por ello se recomienda a los países latinoamericanos pensar holísticamente, o sea, *“pensar un sistema de evaluación que cubra distintas funciones con distintos instrumentos, teniendo siempre presente la cultura del país y su capacidad técnica instalada”*. Para Costa Rica, las buenas prácticas recomendadas serían *“empezar en edades tempranas del estudiantado con evaluación formativa, diagnóstica, y aplicar evaluaciones certificativas en edades más altas”*. (E: Santiago, 2022)

Aquí es necesario un repaso histórico de las macro-evaluaciones al final del ciclo de educación obligatoria para entender la situación actual de Costa Rica y la relevancia de examinarlas en términos de buenas prácticas para retroalimentar el futuro del país en este sentido. Las evaluaciones de certificación han sido muy discutidas y objetadas en años recientes en América Latina, al punto que a la fecha solo las aplican República Dominicana (Ministerio de Educación, 2016) y Colombia (ver sección Gobernanza y organización institucional, apartado Colombia). Costa Rica la eliminó recientemente. Nos referimos al *Examen Nacional de Bachillerato*, que existía en el país reglamentado desde 1892 (primero con exámenes de grado orales con tribunales) y luego se estableció para el último año del bachillerato secundario en 1960. Tuvo una suspensión entre 1974 y 1987 hasta que se reinstaló en 1988, momento desde el cual continuó ininterrumpidamente hasta ser derogado por el Consejo Superior de Educación, CSE, en 2019. (Ruiz, 2018)

La razón alegada para eliminar este examen fue el alto número de estudiantes que venían acumulándose durante años habiendo cursado su secundario, pero sin poder ir a la universidad u obtener un buen empleo por no haber aprobado el examen de Bachillerato después de varios intentos. En 2018 el MEP estimaba unos 27.000 de ellos acumulados en los anteriores 5 años (Semanario Universidad, 10/7/18). Nótese que desde su derogación hasta principios de 2023 no fue reemplazado como requisito final de aprobación y certificación de la educación secundaria. Las autoridades del MEP y el CSE optaron por sustituirlo por un examen muy distinto que, como se apuntó en la sección pasada, tampoco sobrevivió mucho.

El *Examen Nacional de Bachillerato* fue sustituido por otro aplicado por primera (y única vez) en 2021: la prueba *FARO, Fortalecimiento de Aprendizajes para la Renovación de Oportunidades*. La nueva prueba, de objetivos y diseño diferentes, se propuso como una *prueba diagnóstica* por realizar en el penúltimo año previo a la finalización de un ciclo escolar (5° de primaria y 11° de secundaria) en las asignaturas de Matemáticas, Español, Ciencias y Estudios Sociales. No se centraban en contenidos, sino en las habilidades y competencias establecidas por la nueva política curricular que rige desde 2015. Si las pruebas no se aprobaban, se podían repetir y su puntaje representaría el 40% de las calificaciones del año siguiente del estudiante, el cual se promediaba con las calificaciones regulares de ese año. Debían rendirse, pero no eran obstáculo para graduarse y obtener el título.

Debido a una serie de errores en su implementación la única vez que se aplicaron y a las críticas posteriores, las pruebas FARO se suspendieron –primero como promesa de campaña de la administración que asumió el 8 de mayo de 2022 y después, resuelta oficialmente por el CSE. Desde ese momento y durante todo 2022, el sistema educativo de Costa Rica careció de macro-evaluación estudiantil. Cabe aquí traer a la reflexión una certera observación de Ravela et al.: *“En este terreno, al igual que en el económico, se requiere de cierta estabilidad en las reglas de juego. Si la política de evaluación cambia permanentemente, se genera desconfianza y se pierde credibilidad.”*(2008, p.17)

No obstante, siguió activa en el país la discusión sobre si derogar la prueba de certificación de estudios secundarios en 2019 fue o no una buena práctica. Muchas voces expertas se manifestaron en contra de la medida, pidiendo su restablecimiento. (E: Pacheco, 2023; Ruiz,

2022) Argumentan que examinar las prácticas de los países económicamente más desarrollados de Occidente, que tienen los sistemas educativos de más alto desempeño estudiantil del mundo a juzgar por sus resultados en las pruebas PISA, nos muestra que no contar con ningún examen de este tipo al final del ciclo secundario no es una buena práctica, sino lo contrario. También observan que todos los países latinoamericanos (que excepto Colombia y República Dominicana carecen de esta evaluación) califican regularmente en el último tercio de los países del mundo participantes de PISA. Mientras tanto, los países europeos que mejor se desempeñan en PISA toman pruebas académicas rigurosas de salida de la educación secundaria. (E: Ruiz, 2018) Ellas sirven para ingresar a la universidad y además califican para la obtención de buenos puestos de trabajo si el estudiante escoge una ruta posterior de salida laboral, no académica. Pueden ser escritas, o escritas y orales, según decisión de los países; se califican por estratos o *niveles de desempeño* (por ej. excelente, muy bien, bien, suficiente, reprobado) e incluyen ítems de selección tanto como de desarrollo.

En el caso costarricense, finalmente la cuestión se resolvió a inicios del año 2023. El 30 de enero de este año, el CSE acordó por unanimidad y en firme reformar el Reglamento de Evaluación de los Aprendizajes introduciendo una *Prueba Nacional Estandarizada* para el 6º año de la Educación General Básica (educación primaria) y el 11º año de la educación media, en las asignaturas Español, Matemáticas, Ciencias, Idiomas, Estudios Sociales, y Educación Cívica (en secundaria). Será censal, anual, y es requisito para obtener el Certificado de conclusión de la Educación General Básica y el Título de Bachiller en Educación Media. Tendrá dos aplicaciones, una diagnóstica en el primer semestre del curso lectivo y otra sumativa, en el segundo semestre. (CSE, 2023)

Volviendo a las macro-evaluaciones de certificación en general, ellas, como otras, tienen sus propias particularidades en cada país en cuanto a diseño, quién las ejecuta, énfasis en ciertas mediciones, tipos de ítems, etc. Con fines ilustrativos, veamos algunos ejemplos.

En cuanto a quién la ejecuta, algunos países prefieren una evaluación externa a las escuelas (realizada por una agencia nacional independiente como las descritas en la sección anterior o por una dependencia del ministerio que goce de relativa autonomía técnica) porque tiene la ventaja de una mayor confiabilidad y asegura, según las investigaciones, que los estudiantes alcanzan un mejor desempeño en la prueba porque ponen un mayor esfuerzo en prepararse. Otros países favorecen una evaluación interna de las escuelas (realizada por los profesores en sus clases) porque consideran que así los docentes pueden hacer una valoración más completa de las habilidades de sus estudiantes. Respecto a este punto, la mayoría de los países miembros de la OECD confían en un sistema mixto, que combina evaluaciones continuadas realizadas por los docentes durante la secundaria y un examen al final del ciclo que se diseña y procesa externamente. (Dufaux, 2012, p.20)

Pruebas de certificación que se aplican en Europa son el “Baccalaurate” o Prueba de Bachillerato en **Francia**; la Prueba de Matriculación en **Finlandia**, y las pruebas de “Maturitá” (Maduración) en **Italia, Austria, Hungría, Suiza, Bosnia y Herzegovina, Bulgaria, Croacia, República Checa, Albania, Kosovo, Liechtenstein, Macedonia, Montenegro, Polonia, Serbia, Eslovaquia y Ucrania**. Estas pruebas pueden tener énfasis distintos, algunas más orientadas a medir el

desempeño estudiantil (como en **Francia** y **Finlandia**) y otras dirigidas a retroalimentar al sistema (por ej. **Austria**, **Hungría** y **Reino Unido**). Los países además toman en cuenta de diferente manera los promedios de calificación obtenidos por los estudiantes en sus colegios. (Ruiz, 2018)

En los escasos países europeos donde no se toman pruebas para aprobar y certificar la educación secundaria existen pruebas de ingreso a las universidades para quienes deseen continuar con la educación superior. Este camino predomina en los sistemas educativos de Asia. Allí no hay pruebas de salida de secundaria, pero sí hay pruebas de entrada a las universidades que son muy fuertes, examinan una cantidad de materias y se orientan a medir habilidades y capacidades de los aspirantes. Son ej. **Japón**, **Corea del Sur** y **China**. En síntesis, “*en los sistemas educativos de calidad hay siempre al menos uno y a veces dos instrumentos de evaluación (asociados a promoción o avance) en la salida de la secundaria para potenciar la calidad educativa*” (Ruiz, 2018).

¿Cuáles son, entonces, las funciones de la prueba de certificación del bachillerato que hay que valorizar? Siguiendo a Ruiz (2018), coincidimos con que:

- Brindan una clausura o cierre a una etapa de la preparación estudiantil adolescente, que ha llegado a su madurez (“maturitá”), o sea, cierran un proceso formativo y abren nuevos desafíos;
- Ofrecen una garantía a la sociedad sobre la labor que cumplió el sistema educativo preuniversitario;
- Apoyan la cultura evaluativa del país, algo que la OCDE considera una tendencia necesaria y productiva en la educación;
- Son un instrumento clave de persuasión e influencia sobre el personal docente y todo el entorno escolar para que se implemente el currículo oficial y sus orientaciones sobre prácticas educativas, y
- Son un tamiz para la educación superior, que es importante en un contexto nacional donde abundan las universidades privadas con bajos estándares.

En la misma línea, Ravela et al. refuerzan e ilustran el razonamiento:

Un sistema de evaluación para certificar los aprendizajes de los estudiantes a través de exámenes nacionales de alta calidad técnica tiene ventajas asociadas a la transparencia y responsabilidad por los resultados. Hoy es común que dos estudiantes que han aprobado un mismo nivel educativo en dos regiones distintas de un país, posean en realidad niveles de conocimiento muy diferentes. Los sistemas de evaluación para la certificación dan transparencia al valor de los certificados educativos ante la sociedad. Además, dichos sistemas hacen responsables tanto a los docentes como a los propios estudiantes por alcanzar los conocimientos y capacidades que evalúa el examen, lo cual tiene impactos positivos sobre el aprendizaje. En todo caso, la evaluación externa con propósitos de certificación es más apropiada para los ciclos superiores del sistema educativo, sobre todo hacia el final de la educación media. (2008, p. 8)

A la luz de los anteriores argumentos, que ponen el acento en la calidad educativa, debemos plantearnos al mismo tiempo y con igual empeño, ¿cómo se conjugan la calidad y la equidad social en el sistema educativo? Sin duda alguna, la equidad es un desafío real para nuestras sociedades, en especial las latinoamericanas, que están entre las más inequitativas del mundo. Nótese que:

Si las pruebas son exigentes, ello puede conducir a niveles de reprobación muy altos, que afectarían principalmente a los sectores sociales más vulnerables y podrían agudizar problemas como la deserción y el desempleo juvenil, lo cual las haría insostenibles desde el punto de vista social y político. Por ello, es preciso acompañar este tipo de evaluaciones (en realidad todo tipo de evaluaciones, pero estas en particular) con acciones pedagógicas de carácter complementario y compensatorio. Proponer estándares altos implica asumir responsabilidad sistémica por proporcionar todas las oportunidades para aprender: textos, materiales, instalaciones y preparación de maestros, entre otros. (Ravela et al, 2008, p. 52)

Por eso es que hay que contemplar y resolver problemas muy concretos como los que bien resume Ruiz (2018):

- La no retención en la educación secundaria de alrededor de la mitad de estudiantes que ingresan a ella es un problema serio, que afecta a una población numerosa y en su mayoría procedente de los sectores socio-económicos más desfavorecidos;
- La mala calidad de los egresados es un problema nacional que conspira contra la equidad social porque, en general, las instituciones tanto públicas como privadas no están generando las competencias básicas para el Siglo XXI (según los resultados de pruebas comparativas internacionales) y esto no aporta a que muchos jóvenes rompan con el peso de sus condicionantes socio-económicos de origen; y
- La calidad desigual y la ineficiencia de la gestión en muchos centros educativos afecta al sistema en distintos sentidos, por ej. en brindar una limitada formación inicial y en servicio del personal docente; en las brechas que se crean entre colegios públicos y privados, así como entre colegios públicos urbanos y los rurales y nocturnos, y en no incorporar los tiempos de capacitación en la acción docente.

En relación con el asunto que nos ocupa, las macro-evaluaciones de certificación, consideremos cómo se puede avanzar en la equidad sin debilitar la calidad educativa que ellas buscan garantizar, pues ello a la larga ampliaría aún más las brechas sociales. (Ruiz, 2018) Para enfrentar el problema se pueden desarrollar varias líneas de acción. Primero, mejorar las pruebas incorporando ítems de evaluación de habilidades y capacidades superiores que las alinee con la transformación curricular *Educar para una Nueva Ciudadanía* (MEP, 2015) y a la vez flexibilizarlos para ofrecer posibilidades de escogencia a los estudiantes.

Segundo, utilizar mejor las pruebas a fin de que proporcionen algún diagnóstico útil para tomar decisiones a futuro, tal como usarlas para capacitaciones y desarrollo profesional docentes y para tomar decisiones –y no que sean una “autopsia” o “certificado de defunción” que solo mire hacia atrás. (Ruiz, 2018)

En tercer lugar, se pueden diseñar acciones especiales de apoyo a poblaciones que pudieran tener bajo rendimiento. Se podría establecer que la evaluación estandarizada represente solo una parte de la certificación de aprendizajes de los estudiantes (por ej., el 40% o 50% de su calificación final); el resto de la nota sería el de las evaluaciones que realizan los docentes en sus aulas. Esta es una forma de combinar la evaluación externa y la interna, e incluso de aportar a los docentes una mirada desde afuera que los ayude a reflexionar sobre sus propios criterios (Ravela et al., 2008, p. 8). Otro camino que acude a organizar acciones de apoyo es, por ej., el de **Finlandia**, que cuenta con un 30% del estudiantado de bajo rendimiento y lo atiende con acciones de apoyo especial en las áreas de Matemáticas y Lengua.

En cuarto lugar, se puede considerar alguna acción afirmativa especial para poblaciones desfavorecidas socio-económicamente, por ej., valorar la posibilidad de una curva a priori de un puntaje a determinar –digamos, 10 puntos de 100– que se aplicaría a regiones especiales y colegios nocturnos, lo cual haría necesario que la prueba se diseñara técnicamente contemplando esta característica. (Ruiz, 2018)

En síntesis, un sistema educativo tiene a su alcance decisiones y recursos para poner en equilibrio la defensa de la calidad y de la equidad en educación. Solo tiene que proponerse pensarlos y ejecutarlos.

Técnicas e instrumentos de medición

Al enfocarnos en este factor examinamos la metodología de evaluación, es decir los procedimientos, técnicas e instrumentos que se emplean y el valor agregado de usarlos. Entran aquí una cantidad de opciones de recolección y medición de información de las cuales hay que examinar sus características y funcionamiento en relación con los objetivos que persiguen. Es decir, analizar su potencial para alcanzar al máximo posible los resultados deseados evitando o minimizando limitaciones e inconvenientes.

En este sentido, la investigación y la literatura especializada reconocen una cantidad de buenas prácticas metodológicas que contribuyen a construir macro-evaluaciones robustas y fructíferas en aportes de información. Las más destacadas son: (a) *alinean la evaluación estudiantil con la naturaleza del currículo y basar las pruebas de macro-evaluación en los estándares educativos;* (b) *acompañar las pruebas con estudios de factores asociados que contribuyan a explicar los resultados según variables del contexto;* (c) *incorporar a los procesos de evaluación el uso de tecnologías de información y comunicación,* y (d) *abrir la discusión sobre la macro-evaluación nacional a experiencias y aportes internacionales.* Seguidamente se describen, haciendo referencia a países o instituciones especializadas que las emplean.

Alinear la evaluación estudiantil con la naturaleza del currículo y basar las pruebas de macro-evaluación en los estándares educativos.

Un elemento clave del diseño de la evaluación de los aprendizajes es que guarde congruencia con su principal referente: el currículo escolar. Los instrumentos que se utilicen y sus características deben depender de la naturaleza del currículo, porque no puede ser igual la

evaluación que se practique en el marco de un currículo centrado en aprender contenidos que en el de otro basado en desarrollar habilidades o competencias. La evaluación tiene que estar claramente alineada con el currículo, en particular con sus objetivos –o, según un enfoque más actual y extendido, sus *estándares* educativos.

El énfasis contemporáneo en los resultados de aprendizaje subrayó la importancia de contar con *estándares educativos* que aseguren la calidad del trabajo de escuelas y agentes escolares, e impulsó acciones para avanzar hacia esos estándares en los países que no disponen de ellos. 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. En muchos países hay un énfasis creciente en desarrollar y usar estándares educativos ambiciosos como bases de la evaluación y la rendición de cuentas. Al crear un conjunto de estándares contra los cuales medir el desempeño estudiantil, los países buscan evaluar a los estudiantes contra un resultado deseable medible. (OECD, 2013)

Reconocer la utilidad de contar de contar con estándares educativos no debe hacernos descuidar la necesidad de cuidar la coherencia entre currículo, estándares, enseñanza y evaluación. Nusche advierte que este es un desafío para la evaluación:

La lógica central de sistemas basados en estándares se da por el alineamiento de estos elementos clave. Si la evaluación no corresponde al currículo y a los estándares, entonces los resultados de evaluación tienen poco valor para juzgar cuán bien los estudiantes están aprendiendo. Esto, en su momento, volverá difícil el diagnosticar y responder a las necesidades de estudiantes e instituciones educativas. La política entonces necesita darle considerable atención a estrategias sólidas que evalúen el desempeño estudiantil en relación con el currículo y los estándares. (Nusche, 2016, p. 842, citado por Ruiz, 2017, p. 204)

Al establecer estándares nacionales, los resultados estudiantiles se pueden controlar más fácilmente por calidad y son más comparables. (Wang, Beckett y Brown, 2006, citado por OECD, 2013, p.43). El movimiento orientado a comparar los resultados estudiantiles con estándares también ha tenido influencia en motivar a los países a adoptar macro-evaluaciones estandarizadas.

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. Por lo tanto, los estándares educativos varían en especificidad y énfasis entre países (OECD, 2013, p.43).

Acompañar las pruebas con estudios de factores asociados que contribuyan a explicar los resultados según variables del contexto.

El concepto de “factores asociados” se utiliza para designar a los estudios que se realizan en simultáneo a la aplicación de pruebas de macro-evaluación, por lo general mediante la aplicación de cuestionarios complementarios que recogen información sobre las características socio-económicas y culturales del estudiantado y sus familias, el funcionamiento de los centros escolares a los que asisten y la metodología pedagógica con que se imparten asignaturas específicas, a efectos de identificar qué datos de contexto se asocian con los resultados de las pruebas. En otras palabras, los factores asociados determinan variables que inciden positiva o negativamente en el proceso de enseñanza-aprendizaje. (Ravela, n.f. y Backhoff, Bouzas, González, Andrade, Hernández y Contreras, 2008, citados por Flores Reyes, 2010)

La importancia de tal tipo de estudios radica en que, al identificar factores contextuales que favorecen o por el contrario dañan el aprendizaje estudiantil, permiten que las autoridades educativas los analicen y diseñen políticas públicas para fortalecer aquellos que son favorables y debilitar aquellos desfavorables, tras la meta clave de mejorar los aprendizajes estudiantiles. Esta buena práctica ocupa un lugar central en la metodología de dos macro-evaluaciones internacionales muy reconocidas en América Latina, en las cuales participa Costa Rica.

La primera, de carácter continental, es la *serie Estudio Regional Comparativo y Explicativo, ERCE*, que realiza el Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación, LLECE, de OREALC-UNESCO, con sede en Santiago, Chile. Esta prueba se realizó cuatro veces (1997, 2006, 2013 y 2019) y se aplica en 3° y 6° grados de escolaridad para evaluar las áreas de Lectura, Escritura (desde 2019), Matemática y Ciencias (6° grado). Su enfoca en *lo que los estudiantes saben de cada área* (preguntas de opción múltiple y abiertas) y en *los contextos de aprendizaje* (cuestionarios a cuatro actores educativos con los cuales se investigan factores asociados). Veamos qué incluye cada uno de estos cuestionarios.

- El de *estudiantes* examina sus características personales; sus actitudes y percepciones acerca de sus profesores y sus prácticas; sus percepciones sobre las relaciones con sus compañeros y las actividades que realizan fuera de la escuela.
- El de *familias* indaga las características socioeconómicas, demográficas y culturales del hogar; el comportamiento del niño/a; el involucramiento familiar en el proceso de aprendizaje; las expectativas educativas que los padres tienen con respecto a sus hijos/as y las características del barrio en que se inserta la escuela.
- El de *docentes* pregunta sobre sus características personales y antecedentes laborales; sus prácticas docentes; sus expectativas acerca de sus estudiantes y sobre el clima, el liderazgo y la gestión escolar.
- El de *directores de escuela* explora sus características personales; el entorno e infraestructura de la escuela, y la gestión escolar.

Para hacer más evidente el sentido y aportes de los cuestionarios de factores asociados, obsérvense las preguntas centrales que se busca contestar a través de ellos:

- Estudiantes y sus familias: ¿En qué medida se ven asociaciones entre las características de los estudiantes e indicadores de logro en lenguaje, matemática y ciencias? ¿En qué medida se ven asociaciones entre las características familiares e indicadores de logro en lenguaje, matemática y ciencias?

- Importancia de la escuela: ¿Cuán relevante es la escuela a la que asisten los estudiantes para el logro en lenguaje, matemática y ciencias?
- Desigualdades educativas: ¿En qué medida las condiciones socioeconómicas individuales y a nivel de escuela se relacionan con las diferencias de logro en lenguaje, matemática y ciencias?
- Prácticas docentes y ambientes escolares: ¿En qué medida se ven asociaciones entre los procesos pedagógicos e indicadores de logro en matemática, lenguaje y ciencias?
- Características de las escuelas: ¿En qué medida se ven asociaciones entre las características de las escuelas y los indicadores de logro en matemática, lenguaje y ciencias? (Henríquez Calderón, Conferencia y power point, LLECE-UNESCO, 2022)

La segunda macro-evaluación, de carácter internacional, es el ***Programa para la Evaluación Internacional de Estudiantes - PISA***, que desde 2000 promueve la Organización para la Cooperación y el Desarrollo Económico, OECD, en sus países y territorios miembros y asociados. El universo de estudio de PISA abarca a estudiantes que tienen entre 15 años y 3 meses y 16 años y 2 meses de edad. Considerando que los países difieren en la duración y características de sus sistemas educativos, la población no se define en función del grado escolar que cursa sino de la edad a fin de garantizar que los grupos sean equivalentes y los resultados de todos los países puedan ser comparables. Las pruebas PISA se realizan cada 3 años, con la sola excepción de 2021, año en que no se realizó debido a la pandemia y se pospuso para 2022.

Desde su inicio, PISA evaluó tres materias principales: Lectura, Matemática y Ciencias, priorizando una de ellas en cada aplicación consecutiva. Utiliza un enfoque que no está centrado en que el estudiante reproduzca contenidos curriculares sino en que demuestre sus *competencias*, entendidas como la capacidad de aplicar sus conocimientos y habilidades a determinadas áreas disciplinarias y de analizar, razonar y comunicarse con eficacia cuando plantean, interpretan y resuelven problemas en diversas situaciones de la vida real. (Rodino, 2010)

PISA tiene una fuerte orientación hacia impulsar políticas educativas de mejoramiento de los sistemas escolares, por lo cual conecta los resultados de las pruebas de los estudiantes con datos extra-escolares y escolares, por ejemplo, sobre sus características familiares, socio-económicas, culturales, prácticas de estudio y sobre otros factores clave que inciden en su aprendizaje, tales como la formación docente, su metodología, su compromiso, sus creencias y su estabilidad laboral, más las principales características del centro escolar. Este es su estudio de factores asociados que, tal como ocurre en las pruebas ERCE, se realiza a partir de 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.

El examen de factores asociados en cada país y cada prueba PISA ofrece información de gran valor –actual y longitudinal– para que el país profundice la comprensión del desempeño estudiantil, del proceso educativo en general y de las brechas que lo perturban, a fin de ratificar o rectificar el rumbo de sus políticas y estrategias de calidad y equidad. (Ministerio de Educación y Formación Profesional de España, inee blog, 2017).

El Estado de la Educación dedicó varias investigaciones al tema (ver Informes IV, PEN, 2013; V, PEN, 2015, y PEN, VI, 2017) y en 2016 conformó un equipo con el MEP para analizar los principales factores que se identificaron para Costa Rica en PISA 2015. Las conclusiones del equipo confirman la relevancia de la tarea: *“Este conjunto de hallazgos son útiles para que el MEP siga fortaleciendo sus acciones y políticas educativas, de tal manera que a futuro mejoren los rendimientos de los estudiantes, sigan reduciendo las brechas señaladas y la educación pública sea cada vez de mayor calidad, generando en los estudiantes habilidades para la vida, así como mayores oportunidades en la actual sociedad del conocimiento y en el mercado laboral”*. (MEP, Power point, 2017).

Más allá de los estudios de factores asociados que hacen las pruebas internacionales citadas, también una cantidad de países los aplican en sus evaluaciones nacionales. En Latinoamérica están, por ej., **Chile, Colombia y Uruguay**, mientras que en el resto del mundo, entre los miembros de la OECD destacan **Canadá, la Comunidad Flamenca de Bélgica, Suecia y Australia** (OECD, 2013, p. 418 y 462).

Para Costa Rica es pertinente repensar esta buena práctica a la luz de la anulación de las pruebas nacionales de macro-evaluación FARO en 2022, justificada en meras críticas verbales y mediáticas a los cuestionarios de factores asociados usados al aplicar la prueba en 2021, pero sin base en ningún estudio técnico. (La Nación, 19/6/22)

Es cierto que las críticas señalaron algunos serios errores de implementación (longitud exagerada de los cuestionarios, tiempo excesivo que llevó a los estudiantes responderlos, incorporación de preguntas demasiado detalladas y poco acertadas sobre las condiciones socio-económicas del hogar y omisión de solicitar el permiso parental en el caso de estudiantes menores de 13 años). No obstante, la falta de un auténtico estudio que examinara los errores, las críticas, y las razones por las cuales se produjeron condujo, a juicio de esta investigadora, a una decisión intempestiva e infundada que sepultó el trabajo de varios años por diseñar y ejecutar la prueba FARO sin sustituirla de inmediato por una alternativa. Como expresa el dicho popular, *“se arrojó al niño con el agua sucia de la bañera”*. Y así quedamos a cero en este sentido, pues el Acuerdo del CSE que estableció la nueva *Prueba Nacional Estandarizada* en enero de 2023 no hace referencia a acompañarla con un estudio de factores asociados o de contexto, como contemplaba la derogada prueba FARO.

Incorporar a los procesos de evaluación el uso de tecnologías de información y comunicación, TICs

La expansión de la macro-evaluación con pruebas estandarizadas y el manejo de los numerosos datos que genera se han visto muy beneficiadas por la gran capacidad de las TICs. Su uso introduce muchas mejoras en el proceso, por ejemplo, enfoques de evaluación más individualizados; mejor valoración de las habilidades cognitivas, tales como la resolución de problemas; capacidad para hacer rápidamente evaluaciones de gran escala; confiabilidad al hacerlo y costo reducido para administrar la evaluación del desempeño estudiantil, entre otras. Esto contribuye a que su uso sea cada vez más generalizado y eficiente. (OECD, 2013)

En **Costa Rica** contamos con un ejemplo de esta buena práctica cuya su historia y avances son muy ilustrativos. Se trata de un programa iniciado en 2009/10 entre la Escuela de Lenguas Modernas de la Universidad de Costa Rica, ELM-UCR, con el Consejo Nacional de Rectores, CONARE, y el MEP para fortalecer la enseñanza del inglés a través de la certificación de conocimientos del idioma mediante pruebas de evaluación estandarizada, aplicadas primero a docentes de inglés y más tarde a estudiantes de todas las universidades públicas. El programa dio un salto cualitativo y cuantitativo en 2019, cuando se solicitó a la ELM-UCR montar la prueba en computadora y se agregaron como objetivos “*evaluar y certificar a toda la población costarricense*”. (E: Quesada y equipo, 2022)

Se hizo entonces un piloto con maestras de preescolar con una prueba segura y sólida, pero las limitaciones de la plataforma externa que se venía usando llevaron a la ELM-UCR a desarrollar su propia plataforma, con servidores propios, para alojar allí los exámenes y otras herramientas de apoyo. En total se crearon 8 plataformas que actualmente sirven a las 27 regiones del MEP para que sus Direcciones Regionales vean en tiempo real qué docentes o estudiantes están haciendo las pruebas de dominio lingüístico y los resultados se conozcan de inmediato. Más adelante, por problemas informáticos del MEP, la ELM-UCR debió desarrollar tres formatos de prueba –on-line, híbrida y off-line para los colegios que no tenían internet. Son idénticas, sin diferencias en acceso y equidad. Las pruebas se pueden tomar en computadora, tableta y teléfono celular, y sus resultados se conocen a nivel nacional, regional y local de manera rápida y precisa, todos al mismo tiempo. También tienen versiones de exámenes para niños con adecuaciones curriculares, incluso niños superdotados (E: Quesada y equipo, 2022).

La ELM-UCR cuenta con un banco de ítems bajo su custodia y, en este sentido, un diferenciador de calidad de sus pruebas es que son autóctonas, no versiones adaptadas o “*tropicalizadas*” de pruebas hechas en otras partes del mundo. En paralelo a la construcción de pruebas, conduce varios proyectos de investigación y publicaciones, y promueve la formación constante de su equipo de trabajo por medio de posgrados especializados en el exterior. Como próximo avance, están pidiendo al Consejo Superior de Educación, CSE, que el MEP evalúe la *producción oral* de inglés y ya están desarrollando una prueba oral automatizada con inteligencia artificial –pruebas que a nivel internacional son muy costosas, además de que las empresas internacionales no tienen la logística para cubrir las necesidades de Costa Rica. (E: AQ y equipo, 2022)

En América Latina destaca también **Uruguay**. Las pruebas ARISTAS del Instituto Nacional de Evaluación, INEED, son pruebas estandarizadas muestrales, realizadas íntegramente por computadora. Ello es posible gracias a que el país dedicó muchos años –priorizando razones de equidad– a asegurar conectividad y acceso a herramientas digitales para el estudiantado y el magisterio, con apoyos para su uso. Hace más de una década que implementa el *Plan Ceibal*, iniciativa gubernamental pionera en entregar una laptop a cada estudiante del sistema público, el cual a la fecha ha desarrollado una plataforma de aprendizaje digital efectiva. (<https://www.ceibal.edu.uy/es>)

Por ello, en el presente los resultados de evaluación llegan rápidamente a tomadores de decisiones educativas, los medios y la opinión pública. El sistema cuenta además con una muy buena herramienta –ARISTAS EN CLASE– la cual es una plataforma digital donde se van

acumulando datos a lo largo del tiempo, que los docentes pueden consultar en cualquier momento para usarlos. (E: Severin, 2022) Cuando un docente entra, puede generar una prueba para sus estudiantes y, después de que la aplica, puede comparar los resultados con pruebas nacionales de estudiantes del mismo nivel socioeconómico. Estos resultados vienen con recomendaciones a los docentes. (E: Lasida, 2022)

Actualmente se está empezando a promover que esta herramienta se use en formación docente. La herramienta fue elaborada en el país, por funcionarios del propio INEED, y no presenta problemas con el uso de los datos porque la última ley de gobernanza del INEED eliminó la privacidad de los datos sobre estudiantes, docentes y escuelas para investigaciones educativas. Respecto a las nuevas tecnologías digitales en evaluación, ellas sin duda tienen la capacidad de personalizar los aprendizajes y las necesidades de los alumnos, pero a la vez hay que reconocer que su aprovechamiento depende mucho de la capacidad docente para usarlas en su aula. (E: Santiago, 2022)

Colombia también hace un buen uso de las tecnologías de información y comunicación para llegar a sus docentes y escuelas. Son políticas del ICFES que (a) todos los resultados se suban a una plataforma informática de manera rápida porque de muy poco sirve difundirlos 4 meses después, y (b) el sentido de la evaluación radica en hacer buen uso de esos resultados para retroalimentar el proceso educativo. “*Tan importante es medir como aplicar resultados*”. Se destaca que hay buen trabajo conjunto entre el ICFES y el Ministerio de Educación porque hay puentes tendidos y se valoriza que el Instituto tiene los equipos altamente especializados que exige la materia. (E: Figueroa, 2022)

En **Dinamarca**, los test nacionales con computadora se implementaron oficialmente en 2010 y son flexibles en cuanto a que los ítems se adaptan a los niveles de habilidad de los estudiantes. Es decir, los ítems de la prueba van siendo seleccionados de manera secuencial de acuerdo con el desempeño de cada estudiante en los ítems previos. Estas eficientes pruebas nacionales proporcionan una retroalimentación rápida de resultados a los maestros al día siguiente, lo que facilita mucho el uso docente de los resultados de las pruebas. (Danish Ministry of Education y Rambøll, 2012, citado en OECD, 2013).

En **Noruega**, en 2008 todas las escuelas primarias y secundarias ya estaban usando un sistema de administración de pruebas y de ejecución de pruebas electrónicos. Fue posible aplicar pruebas basadas en TICs a cualquier interesado, incluso a estudiantes que tomaban las pruebas nacionales estandarizadas (Norwegian Directorate for Education and Training, 2011, citado en OECD, 2013). Otros ejemplos incluyen el desarrollo de evaluaciones rápidas –por ej., un enfoque facilitado por computadora de pruebas formativas frecuentes y breves; modelos de valor agregado más sofisticados para determinar la contribución de una escuela o un maestro al aprendizaje de un estudiante, y sistemas de datos que proporcionan nuevas oportunidades de compartir información entre agentes escolares. (OECD, 2013, p.45)

Abrir la discusión sobre la macro-evaluación nacional a experiencias y aportes internacionales

Los debates educativos nacionales sobre los resultados estudiantiles están cada vez más influenciados por comparaciones entre países a partir de que se conocen los resultados de pruebas internacionales como PISA, PIRLS, TIMSS y ERCE. Se sostiene que la concepción de la evaluación cambió a lo largo de los años, pasando de una mirada esencialmente nacional, “hacia adentro”, a otra más abierta al exterior, que sin dejar de considerar las aspiraciones y logros educativos del propio país empieza a mostrar interés en compararlos con los de otros países de la región o el mundo. (Bonnet, 2004 en OECD, 2013, p.43)

Algunos sistemas educativos se vieron sacudidos después de conocerse puntajes comparativos internacionales. Por ej., los primeros resultados de PISA publicados en diciembre de 2001 golpearon a la comunidad educativa de **Dinamarca** porque evidenciaron que uno de los sistemas de educación más costosos del mundo estaba funcionando a un nivel que, comparado con resultados de otros países de la OECD, era apenas promedio. Los resultados de PISA contradecían la creencia compartida pero poco sustentada de que las escuelas danesas eran las mejores del mundo. (OECD, 2008)

La creciente disponibilidad de datos internacionalmente comparables sobre el aprendizaje estudiantil influye en las discusiones nacionales sobre educación y ha promovido reformas de política educativa en distintos países. Los datos internacionales comparativos instan a los países a desempeñarse a niveles más altos apoyándose en políticas consideradas potencialmente efectivas en países de alto desempeño. Una cantidad de reformas educativas surge con base en los resultados de pruebas internacionales. Por ej., en **Francia**, la introducción del núcleo común de competencias educativas está vinculado con los resultados de sus estudiantes en PISA. (Dos Santos and Rakocevic, 2012, citado en OECD, 2013, p. 44).

La expansión de evaluaciones internacionales también contribuyó significativamente a que algunos países introdujeran pruebas nacionales estandarizadas. Este fue el caso, por ej., en **Dinamarca, Italia, México y Portugal**, donde previamente se ponía muy poco énfasis en medir resultados educativos. En estos países, la evaluación estudiantil estandarizada sólo estaba disponible a través de pruebas internacionales y se consideró necesario desarrollar mediciones alineadas con los objetivos nacionales de aprendizaje. En algunos países las evaluaciones nacionales se desarrollaron con base en la metodología usada por las evaluaciones internacionales, en particular el marco propuesto por PISA. Esto también impactó en el intercambio de experiencias entre países sobre cómo medir resultados de aprendizaje. (OCDE, 2013, p. 44)

Lo anterior no significa que un país necesariamente debe copiar mecánica o acriticamente procedimientos, técnicas o instrumentos de macro-evaluación que usan las pruebas internacionales (sea que participe en ellas o no) o algún otro país. Lo que recomienda la buena práctica es que la comunidad educativa nacional –autoridades, expertos, docentes, instituciones escolares y también familias– cultiven la disposición y el interés por conocer otras experiencias que ocurren fuera de las propias fronteras para valorar si pueden aprender de ellas algo que les sea útil, sin prejuicios o resistencias derivadas de su propio desconocimiento.

Difusión y uso de resultados

Este factor alude a las estrategias que se ponen en marcha para asegurar que los resultados de la macro-evaluación se conozcan masivamente y tengan incidencia real en las aulas y escuelas del sistema educativo con el fin último de mejorar los desempeños estudiantiles. Es particularmente importante porque, como han identificado expertos latinoamericanos, entre las principales debilidades presentes en los sistemas de evaluación estandarizada en la región se encuentran la estrategia de difusión y el uso de los resultados. (Ravela et al, 2008, p.6) También nuestros entrevistados de **Chile** (E: Cortázar, 2022) y **Uruguay** (E: Lasida, 2022), sin negar la solidez de las macro-evaluaciones en sus respectivos países, señalaron el problema del escaso o deficiente uso de los resultados de evaluación.

Como se hizo al analizar los otros factores, describiremos buenas prácticas recomendadas por la literatura especializada e indicaremos algunos países o instituciones que las aplican con éxito.

Comunicar con amplitud y claridad los resultados de la macro-evaluación a la comunidad educativa y a la opinión pública para sensibilizarlas sobre el estado de la educación y como un compromiso con la transparencia y la rendición de cuentas

En general, en el mundo viene aumentado la oferta de información pública sobre los resultados de la evaluación. Las razones pueden ser una búsqueda de sensibilización de la población general hacia la educación, un compromiso con la transparencia, o la creciente tendencia a asumir la *rendición de cuentas* como otro propósito de la evaluación (OECD, 2013, p.41), o pueden ser todas ellas al mismo tiempo pues todas son valiosas. Lo cierto es que se difunde mucha más información al respecto. Consiste en publicar los resultados de las pruebas estandarizadas a nivel de las escuelas para que los conozcan las familias, funcionarios gubernamentales, medios de comunicación masiva y otros sectores interesados. Esto no solo sirve el propósito de brindar información al público sobre el desempeño del sistema educativo, sino para que los resultados también puedan ser usados por los interesados para actuar, por ej., en cuanto a la elección familiar de las escuelas.

Informar o reportar es un propósito clave de la evaluación y se está volviendo cada vez más importante, según se refleja en los requisitos que establecen diversos países a varios niveles: a nivel del sistema educativo (por ej. reportes sobre el estado de la educación y bases de datos educativos); a nivel de las escuelas (sitios web con información a nivel de escuela, reportes anuales de las escuelas y reportes de inspecciones a las escuelas), y a nivel del estudiantado

(publicación de resultados de evaluaciones estandarizadas por escuela y reporte de notas a estudiantes y sus padres). El énfasis en los informes públicos sobre los resultados de evaluación de las escuelas es particularmente importante en países como **Australia, Bélgica (Comunidad Flamenca), Dinamarca, Islandia y República Eslovaca**. (OECD, 2013, p. 81)

En **Irlanda** en años recientes ha habido un movimiento hacia una mayor rendición de cuentas. Ejemplos de esta tendencia incluyen la introducción de inspecciones escolares regulares a las escuelas secundarias en 2003, la publicación de los informes de las inspecciones escolares en 2006 y la introducción de pruebas estandarizadas obligatorias en las escuelas primarias en 2007. La *Estrategia Nacional para la Alfabetización Lingüística y Aritmética* esquematiza algunas medidas de rendición de cuentas adicionales tales como el desarrollo de estándares nacionales de logros estudiantiles. La estrategia también demanda que las escuelas ofrezcan a los padres información de evaluación adecuada, significativa y clara sobre el progreso de sus hijos. (Irish Department of Education and Skills, 2012, citado en OECD, 2013, p. 42)

En **Chile**, los resultados de la macro-evaluación (Sistema para Medir la Calidad de la Educación, SIMCE) se publican, alimentan con información el sistema de “vouchers” escolares y han contribuido a colocar a la educación en la agenda pública. En **Australia**, los resultados de NAPLAM (Programa Nacional de Evaluación - Alfabetización Lingüística y Aritmética) se publican sobre las escuelas individuales en el sitio web Mi Escuela, donde el público puede tener acceso a datos de desempeño y otros en todo el territorio del país.

Inglaterra se ha venido reportando el desempeño de sus escuelas desde 1988. El movimiento general hacia una mayor transparencia de los resultados incluye la publicación de los informes de inspección a las escuelas, los reportes anuales de las escuelas y los reportes a nivel de la totalidad del sistema educativo que proporcionaban una valoración del estado de la educación.

Poner los resultados al alcance de los docentes y capacitarlos en su interpretación a fin de que sepan usarlos

La macro-evaluación en última instancia solo tiene sentido si la información que genera se utiliza para orientar a docentes y autoridades de las escuelas sobre qué deben hacer para mejorar el desempeño de los estudiantes, en calidad y en equidad. Pero que sea viable que sus resultados se usen para renovar las prácticas escolares (estrategias, recursos, técnicas y materiales) no depende solo de cuánto y cómo se comunican sino, al mismo tiempo, de cómo son recibidos, comprendidos y aprovechados por el magisterio y los administradores escolares. Docentes y directores de escuela difícilmente pueden inferir de los datos duros de la evaluación cuáles son los cambios que necesitan introducir en su pedagogía para mejorar los resultados estudiantiles. Para ellos supondría hacer toda una investigación, lo cual comúnmente excede su formación y sus posibilidades concretas de tiempo y recursos (E: Cuenca, 2022) Por eso, un área en la cual hay una necesidad creciente de desarrollar capacidades en docentes y administradores escolares es la habilidad para interpretar y usar eficazmente los datos provenientes de las evaluaciones estandarizadas.

Ello se debe a que diagnosticar la fuente de las dificultades estudiantiles e identificar los remedios adecuados para diferentes estudiantes es, en sí, una tarea desafiante. Desarrollar la “alfabetización en evaluación” implica las siguientes competencias: la capacidad para examinar datos estudiantiles y entender su sentido; la habilidad para hacer cambios en la enseñanza y las prácticas escolares a partir de esos datos, y el compromiso para participar activamente en discusiones de evaluación externa. Sin desarrollar la capacidad para entender y aplicar evaluaciones, el resultado puede ser “*una lamentable mezcla de confusión, ingenuidad técnica y consejos equivocados*” (Campbell y Levin, 2008, citado en OECD, 2013, p.28)

En **Ontario, Provincia de Canadá**, desarrollar la “alfabetización en evaluación” es responsabilidad de cada distrito escolar. Se comprobó que los distritos escolares de Ontario que mostraban mejores resultados estudiantiles también identificaban entre sus actividades importantes el desarrollo de “alfabetización en evaluación” a nivel de escuelas y distritos. Estas actividades incluían brindar desarrollo profesional a docentes y directivos sobre análisis de datos y alfabetización en evaluación; establecer expectativas sobre el uso de la información de evaluación de los estudiantes; apoyar a las escuelas para que entiendan y usen los datos, y promover el uso de datos para ser capaz de elaborar una planificación de mejoramiento, fijar metas y dar retroalimentación (Campbell y Levin, 2008, citado en OECD, 2-13).

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 macro-evaluaciones estandarizadas. INVALSI crea un “documento de identidad de la escuela” que resume su desempeño en la macro-evaluación nacional junto con información contextual con referencia a 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, citado en OCDE, 2013).

Usar regularmente los resultados de la evaluación para producir mejoras en las prácticas de aula y, en consecuencia, en el desempeño estudiantil

Evaluar en educación carece de sentido si no conduce a renovar la enseñanza en los salones de clase a fin de mejorar el aprendizaje de los estudiantes. Por lo tanto, uno de los puntos críticos al diseñar un sistema de macro-evaluación es garantizar que este se vincule de manera directa y efectiva con las prácticas de aula. Un ejemplo de posibles canales a través de los cuales la macro-evaluación puede llegar a las aulas es asegurar que las escuelas se involucren en prácticas de auto-evaluación significativa e incrementar la capacidad del personal docente para hacer evaluación formativa de los estudiantes.

Con tal propósito, en algunos países el foco se puso en generar estructuras, procedimientos y recursos definidos a nivel central, pero (sin negar su importancia) a este nivel hay mucha menor articulación con las prácticas de aula. Tal enfoque conduce a mayores capacidades técnicas en el *nivel central*, pero a competencias más limitadas en el *nivel local* para involucrarse en actividades evaluativas. Sin embargo, vincular la macro-evaluación y el aprendizaje en las aulas exige necesariamente establecer roles claros en la implementación de las políticas de evaluación para las *estructuras locales* (por ej. directores y supervisores de escuelas y

autoridades educativas locales). El punto clave es que el cumplimiento pleno de la función de desarrollo o mejoramiento que tiene la macro-evaluación requiere articulación con el nivel local. (OCDE, 2013)

En **Australia**, algunas jurisdicciones educativas desarrollaron sofisticados sistemas de información aparte del nivel central. Un buen ejemplo es el del Estado de South New Gales, que construyó una herramienta para hacer análisis de datos llamada *Caja de Herramientas para la Medición, Evaluación e Información sobre las Escuelas* (SMART por su sigla en inglés). Esta iniciativa permite ayudar a los maestros a enseñar a sus estudiantes, retroalimentar a los agentes escolares, servir como plataforma para publicar material de apoyo a los maestros y mejorar la administración del conocimiento, operar como una red para conectar maestros y escuelas con las mismas preocupaciones, y crear mejor infraestructura de datos para la investigación educativa. El sistema SMART ejemplifica cómo la tecnología digital puede ayudar a usar datos de manera eficaz. Adicionalmente, son comunes en las escuelas australianas los sistemas de administración de datos para rastrear el progreso de estudiantes individuales. Esto significa que se puede seguir el desarrollo de cada estudiante a través del tiempo y que esta información se puede compartir con otros docentes o con la siguiente escuela a la que concurra el estudiante. (DEEWR, 2010, y Santiago, Donaldson, Herman y Shewbridge, 2011, citados en OECD, 2013, p. 102)

En **Noruega**, el desarrollo del *Portal de Escuela* contribuyó a que las escuelas tengan acceso a monitorear la información y el análisis de sus propios resultados. Se trata de una herramienta de internet con información clave de monitoreo educativo, que ofrece resultados de aprendizaje, ambientes de aprendizaje, recursos y datos básicos sobre escuelas. El Portal tiene una parte abierta, accesible al público en general, y otra parte protegida por una contraseña, en la que las escuelas pueden obtener información más detallada y compararse con el promedio nacional. Este enfoque es prometedor para incentivar una manera más sistemática y mejor integrada de usar el análisis de datos para que los centros escolares se autoevalúen y hagan sus planes de mejoramiento. A la vez le brinda al público información sobre los resultados educativos nacionales. (Norwegian Directorate for Education and Training, 2011, citado en OECD, 2013, p.102)

De manera similar, en la **Comunidad Francesa de Bélgica**, el *Portal de la Enseñanza* proporciona al público información extensa sobre objetivos de aprendizaje del estudiantado, indicadores educativos y legislación educativa, así como las herramientas pedagógicas y los instrumentos necesarios para que las escuelas se involucren en evaluaciones externas. (Blondin and Giot, 2011, citado en OECD, 2013, p. 102)

Promover políticas educativas basadas en evidencias de la macro-evaluación

Una de las metas principales de un sistema de evaluación es alimentar las prácticas y políticas educativas con evidencias que surjan de la investigación, incluyendo las que se generan al usar los resultados evaluativos. El objetivo es asegurar que los resultados de evaluación se utilicen en investigación académica y de políticas, lo cual posteriormente nutrirá el desarrollo de nuevas políticas.

El principio de construir políticas basadas en evidencias de evaluación está muy claramente establecido en **Nueva Zelandia**. A nivel nacional hay un fuerte compromiso con reunir evidencia nacional e internacional sobre los factores y prácticas que pueden contribuir a mejorar la enseñanza y el aprendizaje. El ejemplo más destacado es el programa *Síntesis de Mejores Evidencias Reiteradas* (BES, por su sigla en inglés) del Ministerio de Educación, el cual analiza investigaciones sobre factores escolares que tienen efecto positivo en el aprendizaje estudiantil. Las publicaciones son ampliamente utilizadas por decisores de políticas y grupos de interés para alimentar la política y la práctica educativa en el país.

Los investigadores y académicos neozelandeses también contribuyen regularmente a debates sobre las políticas de evaluación, tanto individual como colectivamente a través de grupos asesores, como el Consejo de Nueva Zelandia para la Investigación Educativa y la recientemente creada Academia Neozelandesa de Valoración. La Revista de la Oficina de Educación revisa evidencia nacional e internacional sobre prácticas efectivas para apoyar su propio marco metodológico y de indicadores. Sus indicadores se alimentan de la investigación educativa, en particular la Síntesis de Mejores Evidencias antes mencionada y las evaluaciones de escuelas efectivas de la Revista de la Oficina de Educación. (New Zeland Ministry of Education, 2010, citado en OECD, 2013, p. 107)

En **Dinamarca**, el Consejo para la Evaluación y el Desarrollo de Calidad de la Educación Primaria y Secundaria Inferior, como parte de su mandato de monitorear el nivel académico en la educación obligatoria, encarga estudios de investigación y evaluación. Los últimos con frecuencia incluyen estudios sobre la implementación y el uso de nuevas herramientas de macro-evaluación, por ej. Planes de Estudiantes Individuales e informes municipales de calidad, y han conducido a análisis y pruebas pilotos sobre cómo hacer esas herramientas más relevantes para las necesidades locales. Uno de los diez mayores desafíos identificados por el Consejo para el ciclo 2020 es fortalecer el intercambio sistemático de conocimientos entre las instituciones de investigación y las escuelas a fin de promover el uso del conocimiento por parte de estas últimas para mejorar su enseñanza. El Consejo se orienta a recoger y diseminar resultados de investigación con el propósito de apoyar la creación de políticas para mejorar las escuelas. (Danish Ministry of Education and Rambøll, 2011, citado en OECD, 2013, p. 107).

En la **Comunidad Francesa de Bélgica**, la Comisión de Monitoreo, como parte de su misión, revisa toda la investigación en educación y asegura que los resultados relevantes alimenten el desarrollo de políticas y de prácticas escolares. Además, identifica brechas educativas y propone al gobierno un plan de investigación multianual que establece las prioridades de la investigación educacional, así como los objetivos que deben alcanzar esas investigaciones. (Blondin y Giot, 2011, citado en OECD, 2013, p.107)

En **Austria**, la preocupación por fortalecer políticas basadas en investigación contribuyó a la creación del Instituto Federal para la Investigación Educativa, la Innovación y el Desarrollo del Sistema Escolar Austríaco (BIFIE, en su sigla original). Al Instituto se le asignó la responsabilidad de implementar y analizar los estándares educativos, elaborar un examen centralizado de salida basado en competencias, establecer un sistema regular de monitoreo educativo y preparar informes nacionales de educación. (Specht and Sobanski, 2012, citado en OECD, 2013, p. 108)

Varios otros países de la OECD involucran a instituciones nacionales de investigación en la labor de promover políticas educativas basadas en evidencias. Por ej., en **Luxemburgo**, el Ministerio Nacional de Educación y Capacitación Vocacional vincula a las unidades de investigación de la Universidad de Luxemburgo en la mayoría de sus esfuerzos de reforma, en especial en el área de calidad educativa (ADQS, 2011). En **Corea del Sur**, una serie de institutos variados de investigación apoyan a las autoridades educativas –a los niveles central y locales– y a las escuelas con estudios sobre políticas educativas y análisis de la implementación de las políticas (Kim, Kwangho, Gapsung Kim, Sungsook Kim, Jungmin Kim, Jooah Kim y Jungha Park, 2010, citado en OECD, 2013). De la misma forma, **Francia** acude a una cantidad de instituciones investigativas para que asistan a las autoridades específicas en monitorear el sistema educativo y evaluar sus políticas. (Dos Santos y Rakocevic, 2012, citado en OECD, 2013, p. 108)

Conclusiones y recomendaciones

Dado que el objetivo de este trabajo fue identificar las buenas prácticas en materia de macro-evaluación del desempeño estudiantil en la literatura y en las experiencias nacionales e internacionales, las conclusiones a las que arribó la investigación representan, al mismo tiempo, recomendaciones para las autoridades responsables de las políticas educativas del país. Es así pues el propósito final del trabajo fue ofrecer un conjunto amplio de prácticas exitosas, es decir, de probada eficacia, que contribuyen a mejorar la macro-evaluación educativa. Esto es especialmente muy relevante cuando, al momento de finalizarse el estudio en enero de 2023, el sistema educativo costarricense carece totalmente de procesos de macro-evaluación en cualquiera de sus ciclos y niveles.

Pasemos entonces revista a las principales conclusiones-recomendaciones recogidas.

- Las macro-evaluaciones estandarizadas de los aprendizajes hacen *aportes* muy valiosos al sistema educativo que, sin ser únicos ni absolutos, superan los de otros tipos de evaluaciones. A diferencia de las pruebas construidas en el aula, ellas:
 - Permiten visualizar los resultados educativos del conjunto del estudiantado.
 - Ofrecen información sobre el acceso real al conocimiento y las capacidades que logran los alumnos, más allá de su número de años de estudio.
 - Ayudan a hacer visibles aspectos centrales de la labor educativa que no se aprecian de manera directa (por ej. si los estudiantes aprenden lo que se espera de ellos al terminar cada nivel, cuál es el grado de equidad o inequidad en el logro de esos aprendizajes, cómo evolucionan en el tiempo los logros y la equidad en el acceso al conocimiento de los

distintos grupos sociales, y cómo influyen las prácticas escolares y las condiciones de enseñanza en los aprendizajes).

- Aportan información importante a diversos actores sociales, permitiéndoles así participar en la comunidad educativa según su respectivo rol (autoridades, funcionarios directivos y docentes de centros educativos, padres y madres del estudiantado y la ciudadanía).
- Conducen a discusiones informadas sobre qué aspectos del currículo son exigibles a todos los estudiantes y a definir con claridad qué es lo que todos deberían haber aprendido al final de cada ciclo educativo –llámense competencias, metas de aprendizaje, indicadores de logro, niveles de desempeño o estándares.
- A la par de sus importantísimos aportes, conviene también indicar sus *limitaciones*, a fin de evitar simplificaciones. La macro-evaluación estandarizada de los aprendizajes:
 - i. Aporta información fundamental e indispensable sobre la calidad educativa, pero no es un indicador completo de la misma, ya que hay saberes, valores, actitudes y aprendizajes que no pueden o no deben formar parte de ella.
 - ii. Es un componente esencial de un sistema integral de evaluación educativa, pero no es el único tipo de evaluación relevante (también lo son las evaluaciones de los docentes en las aulas, la evaluación del desempeño docente, la evaluación de los centros educativos, la de las políticas educativas, entre otras).
 - iii. Sólo tendrá efectos positivos sobre la educación si se concibe, percibe y emplea como un mecanismo de responsabilidad pública de todos los actores del quehacer educativo.
 - iv. La evaluación en general es condición necesaria pero no suficiente para mejorar la educación.
- A la vista de que en el momento actual el sistema educativo nacional aún tiene que tomar decisiones técnicas sobre la macro-evaluación, esta es una oportunidad ideal para analizar experiencias valiosas que se llevan adelante en el país y fuera de él que pueden inspirar y facilitar el camino de mejoramiento que Costa Rica tiene por delante.
- La macro-evaluación es un proceso académica y técnicamente complejo, en el cual entran en juego diversos factores o variables. Por eso, identificar y seleccionar buenas prácticas en esta materia no puede plantear propuestas únicas, que cubran todos los factores importantes al mismo tiempo y que sean pertinentes para las principales necesidades y aspiraciones educativas de Costa Rica. Las buenas prácticas nacionales e internacionales recogidas ilustran cómo se pueden manejar eficazmente los factores centrales del proceso, a saber: (a) su forma de gobernanza y organización institucional; (b) sus diferentes objetivos; (c) las técnicas e instrumentos empleados para medir los aprendizajes estudiantiles, y (d) el uso que se hace de sus resultados para mejorar la calidad de los aprendizajes.

Sobre la gobernanza y organización institucional de la macro-evaluación

- La literatura especializada coincide en que la mejor práctica es que un país (o una unidad territorial menor como una provincia o estado, si tiene a su cargo la educación) cree una *entidad o agencia especializada de evaluación* con la función específica de diseñar y conducir

el proceso de macro-evaluación, que sea autónoma de las autoridades educativas, esto es, del Ministerio o Secretaría de Educación nacional, provincial o estadual según corresponda.

- Esta práctica representa un enfoque novedoso, actualmente muy extendido en América Latina y el mundo: las autoridades educativas conservan el liderazgo de fijar la estrategia y la política educativa y de implementar las conclusiones de la evaluación, mientras la agencia de evaluación brinda un juicio independiente sobre el logro de los objetivos educativos. Establecer esta agencia autónoma implica darle independencia política y financiera respecto de las autoridades educativas y dotarla de una presencia importante de especialistas.
- Las ventajas de esta práctica son varias. Primero, le otorga a la macro-evaluación cierto grado de independencia frente a las autoridades educativas (que, de otra manera, actúan como “juez y parte”) y fortalece la rendición de cuentas pública de esas autoridades respecto a los resultados del sistema educativo. Segundo, permite contar con una concentración de experticia especializada en la materia, indispensable para desarrollarla de manera adecuada y precisa. Tercero, apoyada en la ventaja anterior, abre la posibilidad de construir una sólida capacidad institucional para aplicar políticas de macro-evaluación. Cuarto, independizar la evaluación educativa de los ministerios de educación puede evitar o reducir vaivenes abruptos en la materia por razones políticas o ideológicas en lugar de razones técnicas, derivadas de cambios en la administración gubernamental.
- La OECD, en su amplio informe de 2017 sobre la educación en Costa Rica, recomienda al país justamente la creación de una agencia independiente de evaluación. Recuerda que la idea no es inédita ya que Costa Rica creó un órgano de este tipo –el Centro Nacional de Evaluación de la Educación, CENE-EDU–, desconcentrado del Consejo Superior de Educación, en 1997 (Decreto Ejecutivo 26014 del 12/5/97)–, el cual desafortunadamente fue eliminado un año después.

Sobre los objetivos de la macro-evaluación

- Los objetivos de la evaluación educativa en general pueden ser (a) *formativos*, dirigidos a retroalimentar los aprendizajes para mejorarlos, y dentro de ellos, como un subtipo, *diagnósticos*, o (b) *sumativos*, orientados a calificar los aprendizajes, dentro de los cuales se incluyen los *objetivos de certificación o acreditación* de aprendizajes adquiridos en el ciclo o nivel que se evalúa. Las instituciones especializadas como la OECD favorecen equilibrar los objetivos formativos y sumativos y recomiendan un sistema que cumpla todas las funciones de la evaluación. Se trata de tener al mismo tiempo dos tipos de evaluación con sus propios instrumentos: una para identificar mejoras necesarias en la enseñanza (por lo común, interna de las escuelas) y otra para calificar y rendir cuenta de los aprendizajes alcanzados (externa a las escuelas y estandarizada, de alcance nacional).
- Ambas evaluaciones deben relacionarse. Las buenas prácticas recomendadas por la OECD para Costa Rica son empezar a edades tempranas del estudiantado con evaluación formativa, diagnóstica, y aplicar evaluación certificativa a edades más altas.

- A la fecha, febrero de 2023, el país aplica evaluaciones formativas y sumativas parciales, que se realizan internamente en las escuelas, y está próximo a aplicar una nueva *Prueba Nacional Estandarizada*, de certificación. (CSE, 2023). El tema de la macro-evaluación de certificación ha sido motivo de debate nacional, aunque la literatura y muchas experiencias internacionales muestran que se puede aplicar salvaguardando la equidad sin rechazar o debilitar la calidad que ella busca garantizar.

Sobre las técnicas e instrumentos de medición de la macro-evaluación

La investigación y la literatura especializada recogen una cantidad de *buenas prácticas metodológicas* que contribuyen a construir macro-evaluaciones robustas y fructíferas en aportes de información. Esta investigación destaca las citadas a continuación.

- Alinear la evaluación estudiantil con la naturaleza del currículo y basar las pruebas de macro-evaluación en los estándares educativos.

Un elemento clave del diseño de la evaluación de los aprendizajes es que guarde congruencia con su principal referente: el currículo escolar. La evaluación tiene que alinearse claramente con él, en particular con sus objetivos –o, según un enfoque más actual y extendido, sus *estándares* educativos. Al establecer estándares nacionales, los resultados estudiantiles se pueden controlar más fácilmente por calidad y son más comparables.

- Acompañar las pruebas con estudios de factores asociados que contribuyan a explicar los resultados según variables del contexto.

Los factores asociados determinan variables que inciden positiva o negativamente en el proceso de enseñanza-aprendizaje. La importancia de estudiarlos radica en que, al identificar factores contextuales que favorecen o por el contrario dañan el aprendizaje estudiantil, permiten que las autoridades educativas los analicen y diseñen políticas públicas para fortalecer aquellos que son favorables y debilitar aquellos desfavorables, tras la meta final de mejorar los aprendizajes estudiantiles.

- Incorporar a los procesos de evaluación el uso de tecnologías de información y comunicación, TICs.

La expansión de la macro-evaluación con pruebas estandarizadas y el manejo de los numerosos datos que genera se benefician mucho de la gran capacidad de las TICs. Usarlas introduce amplias mejoras en el proceso, por ej. enfoques de evaluación más individualizados; mejor valoración de las habilidades cognitivas, tales como la resolución de problemas; capacidad para procesar rápidamente evaluaciones de gran escala; confiabilidad al hacerlo y costo reducido para administrar la evaluación del desempeño estudiantil, entre otras. Esto hace que su uso sea cada vez más generalizado y eficiente.

- Abrir la discusión sobre la macro-evaluación nacional a experiencias y aportes internacionales.

Desde que se conocen los resultados de pruebas internacionales, los debates nacionales sobre resultados estudiantiles están cada vez más influenciados por comparaciones entre países. La concepción de la evaluación cambió en el tiempo, desde una mirada estrictamente nacional a otra más abierta al exterior, que sin dejar de considerar las aspiraciones y logros educativos del país empieza a mostrar interés en compararlos con otros de la región o el mundo. La creciente disponibilidad de datos internacionalmente comparables sobre el aprendizaje influye en las discusiones nacionales y provocó en muchos países reformas de política educativa. Los datos internacionales comparativos instan a los países a desempeñarse a niveles más altos, apoyándose en políticas consideradas potencialmente efectivas en países de alto desempeño.

Sobre la difusión y uso de resultados de la macro-evaluación

Este factor alude a las estrategias utilizadas para asegurar que los resultados de la evaluación se conozcan ampliamente e incidan en las aulas con el fin de mejorar los desempeños estudiantiles. Es muy importante porque los expertos latinoamericanos han identificado que una debilidad considerable de los sistemas de evaluación estandarizada en nuestra región radica en la estrategia de difusión y el uso de los resultados. Las prácticas más recomendadas en este sentido son las siguientes.

- Comunicar con amplitud y claridad los resultados de la macro-evaluación a la comunidad educativa y a la opinión pública para sensibilizarlas sobre el estado de la educación y como un compromiso con la transparencia y la rendición de cuentas.

Se trata de publicar los resultados de las pruebas estandarizadas a nivel de las escuelas para que los conozcan familias, funcionarios gubernamentales, medios de comunicación masiva y otros sectores interesados. Esto no solo sirve el propósito de brindar información al público sobre el desempeño del sistema educativo, sino para que los resultados también puedan ser usados por los interesados para actuar, por ejemplo, en cuanto a la elección familiar de las escuelas.

- Poner los resultados al alcance de los docentes y capacitarlos en su interpretación a fin de que sepan usarlos.

La evaluación solo tiene sentido si la información que genera se usa para orientar a docentes y autoridades escolares sobre qué deben hacer para mejorar el desempeño de los estudiantes, en calidad y equidad. Pero que sus resultados se usen para renovar las prácticas escolares (estrategias, recursos, técnicas y materiales) no depende solo de cuánto y cómo se comunican sino, también, de cómo son recibidos, comprendidos y aprovechados por el magisterio y los directivos. Por eso hay necesidad creciente de cultivar capacidades en tales funcionarios para interpretar y usar eficazmente los datos provenientes de las evaluaciones estandarizadas, o sea, para desarrollar su *alfabetización en evaluación*.

- Usar regularmente los resultados de la evaluación para producir mejoras en las prácticas de aula y, en consecuencia, en el desempeño estudiantil.

Un punto crítico del sistema de macro-evaluación es garantizar que se vincule de manera directa y efectiva con las prácticas de aula. Un ejemplo de posibles canales a través de los cuales la macro-evaluación puede llegar a las aulas es asegurar que las escuelas se involucren en procesos de auto-evaluación significativa e incrementar la capacidad del personal docente para hacer evaluación formativa del estudiantado. Con tal fin, en algunos países el foco se puso en generar estructuras, procedimientos y recursos definidos a nivel central, pero, sin negar su importancia, a este nivel hay mucha menos articulación con las prácticas de aula. Vincular la macro-evaluación y el aprendizaje áulico exige establecer roles claros al implementar las políticas de evaluación para las *estructuras locales* (directores y supervisores de escuelas y autoridades educativas locales). El punto clave es que cumplir plenamente la función de mejoramiento que tiene la macro-evaluación requiere *articulación con el nivel local*.

- Promover políticas educativas basadas en evidencias de la macro-evaluación.

Una de las metas principales de un sistema de evaluación es alimentar las prácticas y políticas educativas con evidencias que surjan de la investigación, incluyendo las que se generan al usar los resultados evaluativos. El objetivo es asegurar que los resultados de evaluación se utilicen en investigación académica y de políticas, lo cual posteriormente nutrirá el desarrollo de nuevas políticas.

Referencias

ADQS, Agency for the Development of Quality in Schools. 2011. OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Luxembourg, Ministry of National Education and Vocational Training, Luxembourg, www.oecd.org/edu/evaluationpolicy.

Agencia de Calidad de la Educación. 2018. *Nuevo Sistema Nacional de Evaluación de Aprendizajes. La evaluación al servicio de los aprendizajes*. Santiago, Chile.

Agencia de Calidad de la Educación. 2023. *Aspectos claves a considerar en el desarrollo de un sistema nacional público de macro-evaluación de los aprendizajes de los estudiantes: recomendaciones desde el caso de Chile*. Documento preparado para el IX Informe Estado de la Educación, San José: PEN.

Backhoff, E., Bouzas, A., González-Montesinos, Andrade, E. Hernández, E. y Contreras, C. 2008. *Factores asociados al aprendizaje de estudiantes de 3º de primaria en México*. México: Instituto Nacional para la Evaluación de la Educación.

- Blondin, C. and B. Giot. 2011. Cadres d'évaluation en vue d'améliorer les résultats scolaires – Étude thématique de l'OCDE : Rapport de base pour la Communauté française de Belgique, Ministère de la Fédération Wallonie-Bruxelles, www.oecd.org/edu/evaluationpolicy
- Bonnet, G. 2004. "Evaluation of education in the European Union: Policy and methodology", *Assessment in Education*, Vol. 11, No. 2.
- Chavarría Navas, S. 2010. *La Política Educativa hacia el Siglo XXI: Propuesta para una Política de Estado*. Ponencia para el III Informe Estado de la Educación, San José: PEN.
- Campbell, C. y B. Levin. 2008. "Using data to support educational improvement", *Educational Assessment, Evaluation and Accountability*, No. 21, pp. 47-65.
- Consejo Superior de Educación, Costa Rica. 2023. AC-CSE-19-03-2023, que aprueba la propuesta de modificación del Decreto ejecutivo No 40862 de fecha 12 de enero del año 2018, denominado Reglamento de Evaluación de los Aprendizajes referente a la Prueba Nacional Estandarizada.
- Contraloría General de la República. 2022. *Informe Auditoría de Carácter Especial sobre el Sistema Nacional de Evaluación de la Calidad de la Educación a cargo del Ministerio de Educación Pública*. Informe N° DFOE-CAP-IF-00009-20022, 8 de Agosto de 2022. San José, Costa Rica.
- Danish Ministry of Education y Rambøll. 2012. OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Update to the Country Background Report for Denmark, National Agency for Quality and Supervision, www.oecd.org/edu/evaluationpolicy
- Danish Ministry of Education y Rambøll. 2011. OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Denmark, www.oecd.org/edu/evaluationpolicy.
- DEEWR, Department of Education, Employment and Workplace Relations. 2010. OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Australia, Canberra, www.oecd.org/edu/evaluationpolicy
- Dos Santos, S. y R. Rakocevic. 2012. Les cadres d'évaluation en vue d'améliorer les résultats scolaires – Étude de l'OCDE : Rapport national de base pour la France, Direction de l'Évaluation, de la Prospective et de la Performance (DEPP), Ministère de l'Éducation nationale, www.oecd.org/edu/evaluationpolicy
- Duffaux, S. 2012. *Assessment for Qualification and Certification in Upper Secondary Education: A Review of Country Practices and Research Evidence*. OECD Education Working Paper No. 83. OECD Publishing, Paris, www.oecd.org/edu/workingpapers

- EPPI. 2002. A Systematic Review of the Impact of Summative Assessment and Tests on Students' Motivation for Learning, Review conducted by the Assessment and Learning Synthesis Group, EPPI, London.
- Francis Salazar, S. 2023. *Análisis de la gobernanza y el proceso de macro-evaluación en Costa Rica: evolución, alcances y desafíos para promover la mejora del sistema educativo*. Ponencia para el IX Informe Estado de la Educación. San José: PEN.
- Ferrer, G. y A. Fiszbein. 2015. *¿Qué ha sucedido con la evaluación de los aprendizajes en América Latina? Lecciones de la última década de experiencia*. Comisión para la Educación de Calidad para todos: documento de antecedentes, El Diálogo. Liderazgo para las Américas y World Bank.
- Flores Reyes, M. G. 2010. *Variables utilizadas para el análisis de factores asociados al rendimiento de los estudiantes*. Ministerio de Educación de Guatemala, Subdirección de Desarrollo de Instrumentos de Evaluación. Guatemala: Dirección General de Evaluación e Investigación Educativa.
- Gronlund, J.L. y Linn, R.L. 1990. *Measurement and evaluation in teaching*. Nueva York: Macmillan.
- Henríquez Calderón, C. 2022. Conferencia y PPP *El desempeño de escolares latinoamericanos en lectura y escritura según el Estudio Regional Comparativo y Explicativo 2019 de LLECE-UNESCO*. Webinar de ProLEER y Estado de la Educación, Costa Rica, 16/9/2022.
- INVALSI. 2010. Sistema di Audit e Accompagnamento delle Scuole PON nella prospettiva dell'Apprendere ad Apprendere – Rapporto Fase 1 Pilota, by R.M. Frascati, <http://archivio.pubblica.istruzione.it/fondistrutturali/documenti/valutazione2007.shtml>.
- Irish Department of Education and Skills. 2012. OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Ireland, www.oecd.org/edu/evaluationpolicy
- Kim, Kwangho, Gapsung Kim, Sungsook Kim, Jungmin Kim, Jooah Kim y Jungha Park 2010. OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Korea, Korean Educational Development Institute (KEDI), Seoul, www.oecd.org/edu/evaluationpolicy
- Lagos, E. 2021. Chile: The Challenge of Providing Relevant Information from ILSA Studies for the Improvement of Educational Quality. En: Nuno Crato (ed), *Improving a Country's Education*, 2021, Springer Nature Switzerland A.G.
- Lasida, J. 2023. *Aspectos claves a considerar en el desarrollo de un sistema nacional público de macro-evaluación de los aprendizajes de los estudiantes: recomendaciones desde el caso uruguayo*. Documento preparado para el IX Informe Estado de la Educación, San José: PEN.

La Nación. 19/6/22. “Pruebas FARO se eliminaron sin análisis técnico”. San José, Costa Rica. <https://www.nacion.com/el-pais/educacion/pruebas-faro-se-eliminaron-sin-analisis-tecnico/MZSM4VMQWZABRIFNTDHNM4E6QA/story/>

Ministerio de Educación Pública, MEP. 2017. *Factores asociados al rendimiento de los estudiantes en la prueba PISA 2015, Costa Rica*. Presentación de power point.

Ministerio de Educación Pública, MEP. 2015. *Fundamentación Pedagógica de la Transformación Curricular. Educar para una nueva ciudadanía*. San José: MEP.

Ministerio de Educación y Formación Profesional de España, Instituto Nacional de Evaluación Educativa, inee blog. 2017. *Análisis de los factores asociados al rendimiento de los alumnos en PISA*. <http://blog.intef.es/inee/2017/02/16>

Ministerio de Educación, República Dominicana. 2016. *Ordenanza No 1-2016 que norma el Sistema de Pruebas Nacionales y de evaluación de los logros de aprendizajes de la República Dominicana*.

Montero, E; Rojas, S., Zamora, E. y Rodino Pierri, A.M. 2012. *Costa Rica en las pruebas PISA 2009 de competencia lectora y alfabetización matemática*. Ponencia preparada para el IV Informe Estado de la Educación. San José: PEN.

New Zealand Ministry of Education (2010), OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for New Zealand, www.oecd.org/edu/evaluationpolicy

Norwegian Directorate for Education and Training. 2011. OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Norway, www.oecd.org/edu/evaluationpolicy

Nusche, D. 2016. Student Assessment and its Relationship with Curriculum, Teaching and Learning in the Twenty-First Century. En: D. Wyse, L. Haward & J. Pandya. *The SAGE Handbook of Curriculum, Pedagogy and Assessment* (pp.838-852). Los Angeles, London, New Delhi, Singapore, Washington D.C: SAGE.

OECD. 2017. *Education in Costa Rica*, Reviews of National Policies for Education, Chapter 3, Basic education in Costa Rica: from access to learning for all. OECD Publishing, Paris.

OECD. 2013. *Synergies for Better Learning. An International Perspective on Evaluation and Assessment*. OECD Reviews of Evaluation and Assessment in Education.

OECD. 2008. “How do OECD countries take stock of progress and performance in education systems? Evidence and issues”, paper prepared for the Education Policy Committee Ad Hoc Workshop “Taking Stock of Education Performance: From Student Testing to System Evaluation”, EDU/EDPC (2008)33, Paris.

- Ospina, M. y Angulo, M.V. 2023. *Aspectos claves a considerar en el desarrollo de un sistema nacional público de macro-evaluación de los aprendizajes de los estudiantes: recomendaciones desde el caso de Colombia*. Documento preparado para el IX Informe Estado de la Educación, San José: PEN.
- Pacheco, F.A. Artículos periodísticos de opinión. “Esta es una gran oportunidad para retomar el bachillerato”, La República, 30 de mayo, 2022. “Una misión de salvamento para la educación nacional”, 18 de junio de 2022. “La aversión a los exámenes”, La Nación, 21 de marzo de 2022. “Nueva vida para el Consejo Superior de Educación”, La Nación, 7 de diciembre, 2021. “El valor del esfuerzo”. La Nación, 17 de mayo, 2018.
- PEN. 2013. *IV Informe Estado de la Educación*. San José: Programa Estado de la Nación.
- PEN. 2015. *V Informe Estado de la Educación*. San José: Programa Estado de la Nación.
- PEN. 2017. *VI Informe Estado de la Educación*. San José: Programa Estado de la Nación.
- Ravela, P. n.f. *Para comprender las evaluaciones educativas*, Fichas didácticas. Chile: PREAL.
- Ravela, P., P.Arregui, G.Valverde, R.Wolfe, G.Ferrer, F.Martínez Rizo, M. Aylwin y L.Wolff. 2008. “Las Evaluaciones Educativas que América Latina Necesita”. *Revista Iberoamericana de Evaluación Educativa*, Vol. 1, Número 1.
- Rodino, A. M. 2013. *El rendimiento de los estudiantes de secundaria en lectura y factores asociados*. Módulo 3, Cuadernos para docentes. San José: PEN. <https://repositorio.conare.ac.cr/handle/20.500.12337/914>
- Rodino, A.M. 2010. *La competencia lectora de los estudiantes costarricenses según la evaluación internacional PISA 2009+*. Ponencia para el IV Informe Estado de la Educación. San José: PEN. <https://repositorio.conare.ac.cr/handle/20.500.12337/858>
- Ruiz, A. 2018. *Pruebas Nacionales de Bachillerato. Perspectivas internacionales*. Power Point presentado ante el Consejo Superior de Educación, CSE.
- Ruiz, A. 2017. *Evaluación y Pruebas Nacionales para un Currículo de Matemáticas que Enfatiza Capacidades Superiores*. México, DF: Comité Interamericano de Educación Matemática.
- Santiago, P., G. Donaldson, J. Herman and C. Shewbridge. 2011. *OECD Reviews of Evaluation and Assessment in Education: Australia*, OECD Publishing, Paris, www.oecd.org/edu/evaluationpolicy.
- Semanario Universidad. 10/7/18. “El país busca cómo graduar a rezagados en bachillerato. San José, Costa Rica. <https://semanariouniversidad.com/pais/mep-busca-como-graduar-a-rezagados-en-bachillerato/>

Specht, W. y F. Sobanski. 2012. OECD Review on Evaluation and Assessment Frameworks for Improving School Outcomes: Country Background Report for Austria, Bundesministerium für Unterricht, Kunst und Kultur (BMUKK), Vienna, www.oecd.org/edu/evaluationpolicy

Wang, L., G. Beckett y L. Brown. 2006. “Controversies of standardized assessment in school accountability reform: A critical synthesis of multidisciplinary research evidence”. In: *Applied Measurement in Education*, Vol. 19, No. 4.

Entrevistas:

Eugenio Severin, Consultor en Educación, Chile, 2022.

Alejandra Cortázar, Agencia de la Calidad de la Educación, Chile, 2022.

Paulo Santiago, División de Asesoramiento e Implementación de Políticas, Dirección de Educación y Habilidades, OECD, 2022.

Javier Lasida, Presidente del Instituto Nacional de Evaluación Educativa, INEED, Uruguay, 2022.

Ricardo Cuenca, Instituto de Estudios Peruanos, Perú, 2022.

María Figueroa, Instituto Colombiano para la Evaluación de la Educación, ICFES, Colombia, 2022

Allen Quesada, Ana Acosta y José Fallas Godínez, Equipo de la Escuela de Lenguas Modernas de la Universidad de Costa Rica, Costa Rica, 2022.

Francisco Miranda López y Deborah Monroy Magaldi, Comisión Nacional para la Mejora Continua de la Educación, México, 2022 y 2023.

Ángel Ruiz, Director del Proyecto *Reforma de la Educación Matemática en Costa Rica*, Costa Rica, 2022.

Francisco Antonio Pacheco, ex Ministro de Educación, Costa Rica, 2023.

María Victoria Angulo y Mónica Ospina, Instituto Colombiano para la Evaluación de la Educación, ICFES, Colombia, 2023.