

Enpath

INFORME · INVESTIGACIÓN

Decisiones públicas basadas en evidencia en América Latina: treinta años construyendo sistemas, un eslabón pendiente

Abril de 2024

Organización independiente dedicada a promover un uso responsable de la inteligencia artificial en las decisiones públicas y organizacionales. Cada publicación incluye su metodología, sus fuentes y sus limitaciones.

América Latina lleva tres décadas construyendo sistemas de seguimiento y evaluación, y algunos son referentes mundiales. Lo que la región no ha construido es el último eslabón: que la evidencia cambie decisiones. Este informe examina esa brecha —de la que depende, además, todo lo que la inteligencia artificial promete al sector público.

Actualización.

Actualizado el: 18/07/2026. Esta sección resume acontecimientos posteriores a la fecha original de publicación y no forma parte del análisis histórico original.

- La serie anual de Enpath *Estado de las decisiones públicas en América Latina* (primeras ediciones: julio de 2025 y julio de 2026) desarrolla y actualiza el diagnóstico de este informe con mediciones propias.
- En México, la reforma administrativa aprobada en 2024–2025 trasladó las funciones del CONEVAL a otras entidades públicas, un cambio institucional mayor para el sistema de evaluación descrito aquí.

Resumen ejecutivo.

La pregunta que este informe responde no es si América Latina produce evidencia para sus decisiones públicas —la produce—, sino si esa evidencia llega a las decisiones. La respuesta corta: **la región institucionalizó la generación de evidencia con más éxito que su uso.**

Los activos son reales y llevan décadas: Colombia creó SINERGIA en 1994, uno de los sistemas de seguimiento y evaluación más antiguos del mundo en operación continua; la evaluación experimental de Progresá en México (1997) fundó el estándar mundial de evaluación de programas sociales; el CONEVAL mexicano, la DIPRES chilena y el presupuesto por resultados peruano completan un mapa institucional que pocas regiones en desarrollo igualan. La medición del BID sobre gestión pública orientada a resultados confirma el patrón: Brasil, Chile, Colombia y México son los únicos países de la región con capacidades consolidadas en ambas mediciones disponibles, con Chile (4,1 sobre 5) y México (4,0) a la cabeza ([BID, 2015](#)).

El pasivo también es real: los sistemas producen informes que el proceso presupuestal no usa, los programas mal evaluados rara vez se rediseñan o terminan, y la alta rotación de equipos disuelve el aprendizaje institucional. Mientras tanto, la confianza ciudadana —la razón última de decidir bien— sigue cayendo: 36,3 % de la población de 16 países de la región confiaba en su gobierno nacional en 2022, frente a 43 % en 2008 ([OCDE, 2024](#)).

Este informe introduce la **escalera Enpath de madurez** —datos, seguimiento, evaluación, uso— para ordenar el diagnóstico, y concluye con recomendaciones centradas en el nivel 4: reglas que conecten evaluación y presupuesto, agendas de aprendizaje por sector y protección institucional de los equipos técnicos. La conexión con la agenda de esta serie es directa: la inteligencia artificial aplicada al gobierno es una promesa de mejores decisiones; sin el eslabón del uso, será un amplificador de la brecha existente.

Hallazgos principales.

1. **La región tiene infraestructura de evidencia de primera generación consolidada.** Cuatro países (Brasil, Chile, Colombia, México) mantienen capacidades de gestión orientada a resultados iguales o superiores a 3,0 sobre 5 en las dos mediciones del BID disponibles; Chile (4,1) y México (4,0) lideran ([BID, 2015](#)).
2. **El eslabón débil es el uso, no la producción.** La OCDE documenta que en la región la evaluación rara vez tiene consecuencias presupuestales o de diseño; los mecanismos que vinculan resultados con decisiones —cláusulas de revisión, agendas de aprendizaje, respuesta obligatoria del evaluado— son la excepción ([OCDE, 2020](#); [OCDE, 2024](#)).
3. **La confianza ciudadana no acompaña la infraestructura técnica.** 36,3 % de confianza en el gobierno nacional (2022) frente a 43 % en 2008; América Latina confía menos en sus gobiernos que las demás regiones comparadas por la OCDE. Producir evidencia sin usarla visiblemente no genera confianza.
4. **Colombia es un caso de liderazgo temprano con agenda incompleta.** SINERGIA (1994) y la práctica de evaluación del DNP son referentes regionales; el punto débil está en el uso presupuestal de las evaluaciones, la continuidad de los equipos y la integración de los registros administrativos.
5. **La evaluación experimental nació como práctica de gobierno en la región.** Progresá (México, 1997) demostró que un programa social podía diseñarse para ser evaluado con asignación aleatoria y sobrevivir cambios de gobierno gracias a su evidencia. Esa lección —evaluar para decidir, no para archivar— sigue siendo la mejor exportación de la región en la materia.

Por qué importa.

Decidir con evidencia no es una preferencia técnica: es la diferencia entre políticas que se corrigen y políticas que se perpetúan por inercia o clientela. Este informe apoya tres decisiones concretas: para los ministerios de hacienda y planeación, **qué mecanismo institucional conecta los resultados de evaluación con el ciclo presupuestal**; para los ministerios sectoriales, **qué preguntas de su agenda deben responderse con evaluación antes de escalar o terminar programas**; y para cualquier entidad que considere adoptar inteligencia artificial, **si su cadena de decisión está preparada para usar la evidencia que esos sistemas producen**. La serie de Enpath sostiene una tesis simple: la IA no arregla cadenas de decisión rotas; las acelera.

Antecedentes.

Tres décadas en tres olas.

Primera ola: sistemas nacionales (1994–2005). Colombia creó SINERGIA en 1994 al amparo de la Constitución de 1991, que ordenó evaluar la gestión pública ([DNP](#)); Chile inició en 1997 la evaluación sistemática de programas desde la DIPRES ([DIPRES](#)); México evaluó Progres a con diseño experimental desde su origen (1997) y creó el CONEVAL por ley en 2004 ([CONEVAL](#)).

Segunda ola: presupuesto y profesionalización (2005–2015). Perú adoptó el presupuesto por resultados (2007); la red regional se profesionalizó con centros como CLEAR LAC (2011) y la expansión de los laboratorios académicos de evaluación experimental. La medición del BID entre 2007 y 2013 registró avances generalizados en planeación y seguimiento, con la brecha mayor en la vinculación entre resultados y presupuesto ([BID, 2015](#)).

Tercera ola: institucionalidad del uso (2015–). Brasil creó en 2016 el comité que coordina la evaluación de políticas federales (CMAP); Estados Unidos convirtió en ley la arquitectura del uso (Evidence Act, 2019: agendas de aprendizaje y responsables de evaluación por agencia); el Reino Unido consolidó los What Works Centres como puente entre evidencia y práctica. Esta tercera ola —la del uso— es la que la región tiene pendiente.

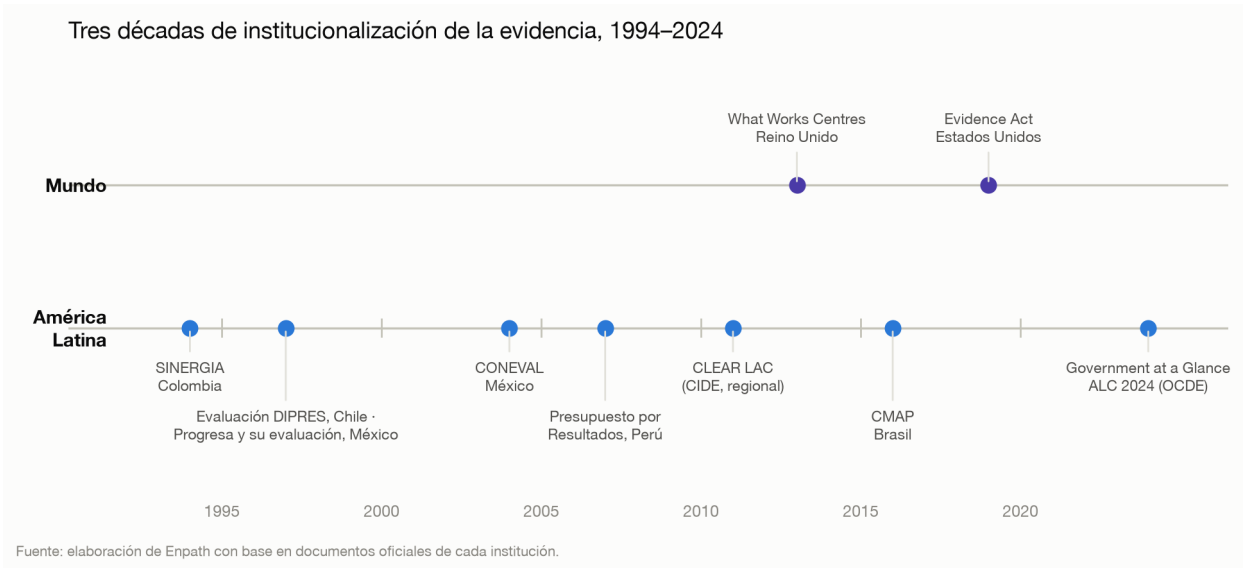
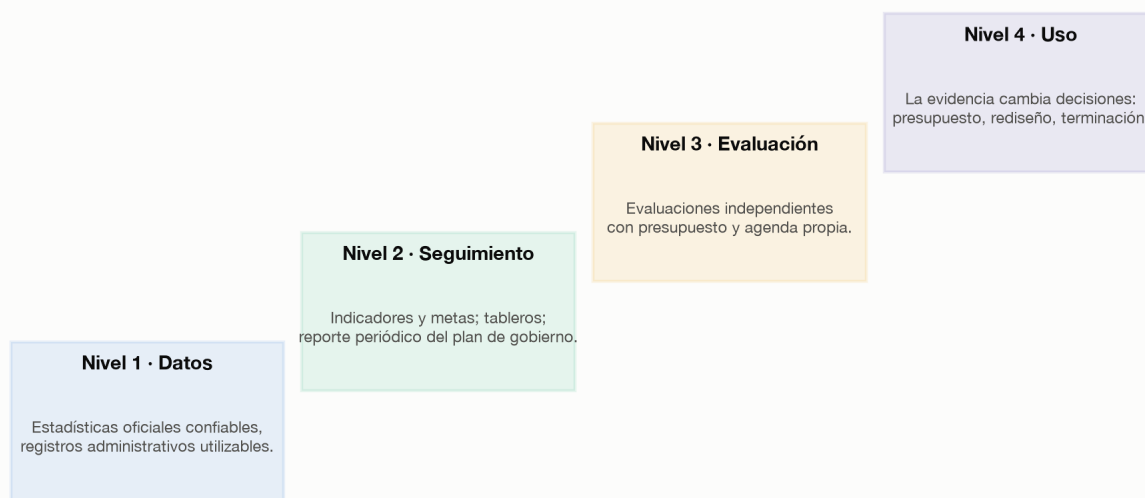


Figura 1. Hitos de la institucionalización de la evidencia, 1994–2024. Fuente: elaboración de Enpath.

Una escalera para ordenar el diagnóstico.

Escalera Enpath de madurez en decisión pública basada en evidencia



La región avanzó con solidez hasta el nivel 3. El eslabón débil es el nivel 4: el uso.

Figura 2. Escalera Enpath de madurez en decisión pública basada en evidencia. Fuente: Enpath.

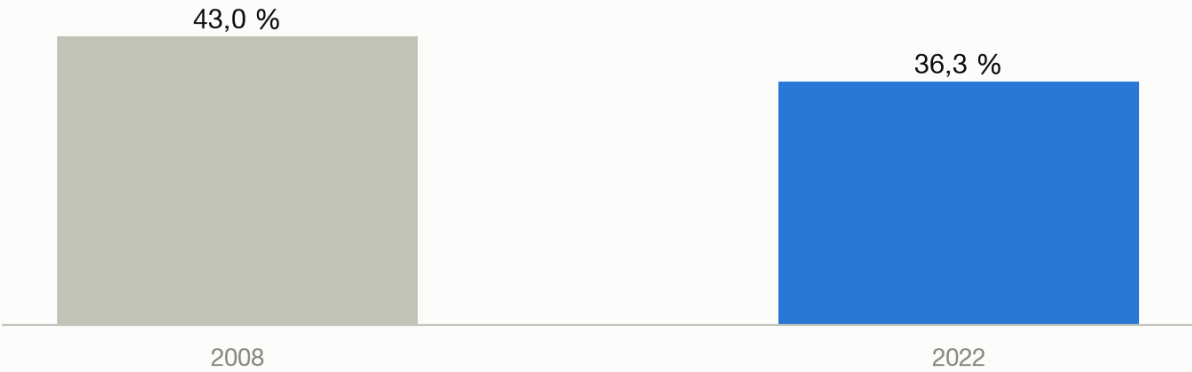
- **Nivel 1 · Datos.** Estadísticas oficiales confiables y registros administrativos utilizables. Sin esto, lo demás es decoración.
- **Nivel 2 · Seguimiento.** Indicadores, metas y reporte periódico del plan de gobierno.
- **Nivel 3 · Evaluación.** Evaluaciones independientes, con presupuesto, métodos públicos y agenda propia.
- **Nivel 4 · Uso.** El nivel que define la madurez real: los resultados cambian presupuestos, rediseñan programas o los terminan; el que decide responde por lo que hace con la evidencia.

El criterio de verificación del nivel 4 es incómodo pero simple: **¿puede la entidad citar tres decisiones de los últimos dos años que cambiaron por resultados de evaluación?** Si la respuesta es un silencio o un ejemplo protocolario, el sistema opera en nivel 3.

Situación actual.

El estado de la región puede resumirse en una asimetría: **madurez alta en los niveles 1 a 3 en un grupo de países, madurez baja en el nivel 4 en casi todos.** Las piezas de la tercera ola —respuesta obligatoria del evaluado, cláusulas de revisión presupuestal, agendas de aprendizaje— existen de forma dispersa y frágil. A la asimetría técnica se suma la política: la confianza en el gobierno nacional bajó de 43 % (2008) a 36,3 % (2022) en 16 países de la región, y la satisfacción ciudadana con los servicios sigue por debajo del promedio OCDE (OCDE, 2024). La evidencia que no se traduce en mejoras visibles no repara esa confianza.

Confianza en el gobierno nacional · América Latina y el Caribe
(porcentaje de la población, 16 países)



Fuente: OCDE, Government at a Glance: Latin America and the Caribbean 2024, con datos de Gallup World Poll.

Figura 3. Confianza en el gobierno nacional en América Latina y el Caribe, 2008 y 2022. Fuente: OCDE (2024), con datos de Gallup World Poll.

Análisis de datos.

Qué se midió y qué muestra.

La medición del BID (24 países, dos olas). El índice de gestión para resultados evaluó cinco pilares — planeación, presupuesto, gestión financiera, gestión de programas, seguimiento y evaluación— en escalas de 0 a 5. Los resultados sitúan a Chile (4,1) y México (4,0) a la cabeza, con Brasil y Colombia completando el grupo de capacidades consolidadas ($\geq 3,0$ en ambas olas). El avance regional se concentró en planeación y seguimiento; el pilar sistemáticamente más débil fue la conexión entre resultados y presupuesto (BID, 2015).

El panorama OCDE (2020 y 2024). El mapeo comparado de la OCDE sobre evaluación de políticas encontró que la mayoría de los países —también fuera de la región— luchan con el uso: la evaluación se produce, pero los mecanismos que obligan a considerarla en decisiones son minoritarios (OCDE, 2020). El *Government at a Glance* regional de 2024 añade el contexto de gobernanza: brechas persistentes en capacidad del servicio civil y en integridad, con la confianza como síntoma agregado (OCDE, 2024).

Sistema	País	Año de origen	Ancla institucional	Producto principal	Talón de Aquiles
SINERGIA	Colombia	1994	DNP	Seguimiento del PND; evaluaciones de política	Uso presupuestal de resultados
Evaluación DIPRES	Chile	1997	Ministerio de Hacienda	Evaluaciones ligadas al ciclo presupuestal	Alcance frente a demanda
CONEVAL	México	2004			

Sistema	País	Año de origen	Ancla institucional	Producto principal	Talón de Aquiles
			Organismo con autonomía técnica	Evaluación social; medición de pobreza	Autonomía bajo presión política
Presupuesto por Resultados	Perú	2007	MEF	Programas presupuestales con indicadores	Calidad frente a formalismo
CMAP	Brasil	2016	Hacienda/Planeación	Evaluación de gasto federal	Consolidación reciente

Tabla 1. Cinco sistemas nacionales de evidencia. Fuente: elaboración de Enpath con base en documentos oficiales de cada institución.

Colombia.

Colombia fue pionera y sigue siendo referente: SINERGIA opera sin interrupción desde 1994, el DNP mantiene una práctica de evaluaciones de política pública con métodos y resultados publicados, y el país cuenta con un sistema estadístico (DANE) que ha modernizado operaciones críticas —censo, encuestas de hogares, estadísticas vitales—. Las evaluaciones de *Familias en Acción* a comienzos de los años 2000, con métodos cuasiexperimentales y efectos documentados en nutrición y asistencia escolar, se convirtieron en un caso de estudio regional sobre cómo la evidencia protege a un programa a través de gobiernos de distinto signo ([DNP–SINERGIA](#)).

El diagnóstico de nivel 4, sin embargo, aplica con fuerza:

- **Evaluación sin consecuencia presupuestal automática.** No existe regla que obligue a considerar los resultados de evaluación en la programación del gasto; la decisión de usar o ignorar una evaluación es discrecional.
- **Rotación que disuelve memoria.** Los equipos técnicos de planeación y evaluación rotan con los ciclos políticos; el conocimiento acumulado se va con las personas.
- **Registros administrativos fragmentados.** La interoperabilidad entre bases (salud, educación, protección social, tributaria) avanza lento, y cada evaluación paga el costo de reconstruir datos que el Estado ya tiene.
- **Asimetría territorial.** La capacidad de seguimiento y evaluación se concentra en el nivel nacional; departamentos y municipios —donde se ejecuta la mayor parte de la política social— operan mayoritariamente en niveles 1 y 2 de la escalera.

América Latina.

El patrón colombiano se repite con variaciones. **México** construyó el sistema de evaluación social más institucionalizado de la región (CONEVAL), cuya medición multidimensional de pobreza es insumo obligado del debate público; su prueba permanente es la autonomía. **Chile** integró la evaluación al presupuesto desde Hacienda —el diseño con mejor conexión producción-uso de la región— con la

limitación inversa: alcance acotado. **Brasil** combina capacidad analítica de primer nivel (IPEA) con una institucionalidad de uso reciente (CMAP). **Perú** demostró que el presupuesto por resultados puede adoptarse a escala, y también que puede volverse formalismo si los indicadores no se auditan. **Uruguay y Costa Rica** mantienen sistemas de seguimiento sólidos a su escala. El resto de la región opera, con excepciones sectoriales, en los niveles 1 y 2.

En paralelo, la región alberga una infraestructura no estatal de primer orden: centros académicos y laboratorios de evaluación experimental que trabajan con gobiernos (la red que creció alrededor de J-PAL, IPA y las universidades de la región), y programas regionales de formación (CLEAR LAC). La materia prima intelectual existe; el cuello de botella es la demanda estatal por sus resultados.

Comparación internacional.

Tres referencias externas iluminan el eslabón faltante:

- **Estados Unidos — Evidence Act (2019).** Convierte el uso en obligación organizacional: cada agencia debe tener una agenda de aprendizaje (qué preguntas necesita responder), un responsable de evaluación y planes anuales públicos. No garantiza que la evidencia gane cada discusión; garantiza que exista un dueño institucional de las preguntas ([Congreso de EE. UU., 2019](#)).
- **Reino Unido — What Works Centres (2013–).** Resuelven el problema de síntesis: traducen la literatura a recomendaciones utilizables por maestros, policías, alcaldes. La lección: el uso requiere intermediarios dedicados, no solo repositorios ([Gobierno del Reino Unido](#)).
- **OCDE — mapeo de evaluación (2020).** La constatación comparada de que el uso es el problema de todos, y de que los países que mejor lo resuelven comparten tres piezas: obligación legal de evaluar, estándares de calidad públicos y mecanismos formales de respuesta a las evaluaciones ([OCDE, 2020](#)).

La comparación deja una advertencia para la región: ninguna de las tres referencias empezó por comprar tecnología. Empezaron por asignar responsabilidad institucional sobre preguntas y respuestas.

Oportunidades.

- **El eslabón del uso es barato en términos fiscales.** Una regla de respuesta obligatoria a evaluaciones, una agenda de aprendizaje por ministerio y cláusulas de revisión en programas nuevos cuestan poco; su efecto es transformar inventario en decisión.
- **Los registros administrativos son la mina sin explotar.** La región produce datos administrativos masivos; integrarlos para evaluación continua (con salvaguardas de privacidad) multiplicaría la evidencia disponible a costo marginal — la agenda del informe de mayo de esta serie.
- **La IA eleva el retorno de la madurez.** Las entidades en niveles 3–4 de la escalera podrán usar sistemas de IA para acelerar síntesis, focalización y monitoreo con provecho; ese retorno es el argumento fiscal para subir de nivel ahora.

- **Ventana de cooperación regional.** La heterogeneidad —Chile fuerte en presupuesto, México en evaluación social, Colombia en seguimiento— es una agenda natural de intercambio, con las instituciones regionales (CEPAL, BID, CAF, CLEAR) como vehículo existente.

Riesgos.

- **Evaluación ritual.** El riesgo maduro de la región: sistemas que producen documentos para cumplir, no para decidir. El síntoma es la evaluación que nadie espera y ninguna decisión cita.
- **Desmonte institucional.** Los sistemas de evidencia son vulnerables a la política: recortes, capturas o supresiones de organismos evaluadores revierten décadas en meses. La independencia técnica requiere protección explícita.
- **Tecnificación sin demanda.** Comprar tableros, lagos de datos o IA sin resolver quién usa las respuestas reproduce el patrón: más oferta de evidencia, igual demanda.
- **Desigualdad territorial creciente.** Si la madurez sigue concentrada en los gobiernos nacionales, la brecha entre política nacional evaluada y gestión territorial a ciegas seguirá creciendo donde más importa: en la ejecución.

Perspectivas.

Con la información disponible a abril de 2024: la OCDE continuará la serie regional (

Government at a Glance

ALC) como línea de base comparada; los procesos de adhesión a la OCDE en curso en la región (Perú, Brasil y Argentina en distintas etapas) presionarán por estándares de gobernanza basada en evidencia; y la agenda de IA pública —con la política nacional colombiana en elaboración— pondrá a prueba la tesis de este informe: los países que lleguen a la IA con el eslabón del uso resuelto obtendrán resultados; los demás automatizarán su statu quo. Enpath planea volver sobre este diagnóstico con una medición anual propia.

Recomendaciones.

1. **Adopte la regla de respuesta obligatoria.** Toda evaluación de un programa público exige respuesta escrita y pública del responsable en un plazo definido: qué acepta, qué rechaza y qué cambiará. Es la intervención de mayor retorno inmediato en el nivel 4.
2. **Instale agendas de aprendizaje por ministerio.** Cinco a diez preguntas prioritarias, publicadas, con responsable y calendario. La evaluación deja de ser un flujo de informes y se vuelve la respuesta a preguntas que alguien espera.
3. **Conecte evaluación y presupuesto con reglas mínimas.** Programas nuevos por encima de un umbral nacen con cláusula de evaluación y fecha de revisión; los resultados acompañan obligatoriamente el anteproyecto de presupuesto del sector.

4. **Proteja la capacidad técnica de la rotación.** Carreras técnicas estables en planeación y evaluación, y documentación institucional de métodos y decisiones: la memoria no puede vivir en los contratistas.
5. **Invierta en registros administrativos interoperables** antes que en nuevas encuestas o plataformas: es la infraestructura del nivel 1 que abarata todos los niveles superiores.
6. **Lleve la escalera al territorio.** Un paquete mínimo para gobiernos locales y regionales — indicadores estándar, datos compartidos, acompañamiento de evaluación— reduce la brecha donde se ejecuta la política.

Metodología.

Revisión de fuentes públicas con corte al 30 de abril de 2024: mediciones comparadas (BID, 2015; OCDE, 2020 y 2024), documentación institucional de los sistemas nacionales citados (DNP–SINERGIA, DIPRES, CONEVAL, MEF, CMAP) y marcos de referencia internacionales (Evidence Act; What Works Network). La escalera de madurez es una elaboración de Enpath para ordenar el diagnóstico; sus niveles se derivan de la literatura comparada citada, no de una medición propia. Las cifras de confianza provienen de la OCDE (2024) con datos de Gallup World Poll.

Limitaciones. No existe una medición regional reciente y comparable de la madurez de uso de evidencia; la medición del BID data de 2013–2015 y el mapeo OCDE de 2020 cubre parcialmente la región. Este informe no realiza medición propia: caracteriza cualitativamente con la mejor evidencia disponible, y esa es su principal limitación —que futuras ediciones de esta serie buscarán resolver.

Referencias.

1. BID — García López, R. y García Moreno, M. (2015, 2.^a ed. de la serie). *Construyendo gobiernos efectivos: logros y retos de la gestión pública para resultados en América Latina y el Caribe*. Banco Interamericano de Desarrollo. <https://publications.iadb.org/publications/spanish/document/Construyendo-gobiernos-efectivos-Logros-y-retos-de-la-gesti%C3%B3n-p%C3%ABblica-para-resultados-en-Am%C3%A9rica-Latina-y-el-Caribe.pdf>
2. Congreso de los Estados Unidos (2019). *Foundations for Evidence-Based Policymaking Act of 2018* (H.R. 4174). <https://www.congress.gov/bill/115th-congress/house-bill/4174>
3. CONEVAL. *Consejo Nacional de Evaluación de la Política de Desarrollo Social* (documentación institucional). <https://www.coneval.org.mx/>
4. DIPRES. *Sistema de Evaluación y Control de Gestión* (documentación institucional). Ministerio de Hacienda de Chile. <https://www.dipres.gob.cl/>
5. DNP. *SINERGIA: Sistema Nacional de Evaluación de Gestión y Resultados* (documentación institucional). Departamento Nacional de Planeación de Colombia. <https://sinergia.dnp.gov.co/>
6. Gobierno del Reino Unido. *The What Works Network* (documentación institucional). <https://www.gov.uk/guidance/what-works-network>
7. OCDE (2020). *Improving Governance with Policy Evaluation: Lessons From Country Experiences*. OECD Publishing. <https://doi.org/10.1787/89b1577d-en>

8. OCDE (2024). *Government at a Glance: Latin America and the Caribbean 2024*. OECD Publishing, 13 de marzo de 2024. https://www.oecd.org/en/publications/government-at-a-glance-latin-america-and-the-caribbean-2024_4abdba16-en.html
9. Skoufias, E. (2005). *PROGRESA and Its Impacts on the Welfare of Rural Households in Mexico*. IFPRI Research Report 139. <https://www.ifpri.org/publication/progres-and-its-impacts-welfare-rural-households-mexico>

Enpath · Investigación aplicada para decisiones públicas basadas en evidencia. Cada publicación incluye su metodología, sus fuentes y sus limitaciones.