

# Evaluación de Política

*Guía paso a paso para diseñar la evaluación de una política pública: pregunta evaluativa, teoría de cambio, indicadores SMART, método y criterios OCDE-DAC*

Esta guía acompaña el uso de la herramienta web en [ricardoruiz.co/evaluacion.html](http://ricardoruiz.co/evaluacion.html). Es el cuarto módulo del Lab de Políticas Públicas y Prospectiva, junto con problema público (Bardach), análisis estructural (MicMac) y análisis de actores (Mactor). Si los tres primeros te ayudan a *diseñar* la política, este te ayuda a *diseñar cómo vas a saber si funcionó* — antes de implementar.

## Contenido

- |    |   |
|----|---|
| 01 | Qué hace el módulo y para qué sirve                               |
| 02 | Mecánica 1 · pregunta evaluativa (tipo + alcance + Sinergia DNP)  |
| 03 | Mecánica 2 · teoría de cambio (marco lógico CEPAL)                |
| 04 | Mecánica 3 · indicadores SMART                                    |
| 05 | Mecánica 4 · selector de método (14 métodos · frontera 2020-2026) |
| 06 | Mecánica 5 · criterios OCDE-DAC                                   |
| 07 | Mecánica 6 · análisis económico (CBA · MVPF · CEA)                |
| 08 | Mecánica 7 · plan operativo + 3 descargas                         |
| 09 | Pre-Analysis Plan exportable (AEA RCT Registry / OSF)             |
| 10 | Copiloto IA y cómo aprovecharlo bien                              |
| 11 | Cómo encadenar con los demás módulos del lab                      |

## 01 · Qué hace el módulo y para qué sirve

Evaluar bien una política pública no es preguntar *¿funcionó?* al final. Es decidir, **antes de implementar**, qué pregunta vas a contestar, qué teoría de cambio estás asumiendo, qué vas a medir y con qué método. El resultado de hacer este trabajo previo es una evaluación defendible incluso si la política falla en alguna parte.

Para qué te sirve, en concreto:

- **Defender el plan ante quien financia.** Cuando puedes explicar por qué pediste tres meses de línea base y un método cuasi-experimental, dejas de ser tratado como burócrata costoso y empiezas a ser tratado como socio del diseño.
- **Blindar el análisis contra el p-hacking.** Si fijas el plan después de ver los datos, vas a ajustar la evaluación a lo que los datos digan. Registrar el plan antes obliga a reportar los hallazgos contra ese plan, no contra una versión revisada.

- **Producir un plan formal exportable.** El entregable es un documento .md estructurado en 6 secciones, compatible con protocolos Sinergia DNP, y una matriz de indicadores en CSV lista para llevar a Excel o cargar en sistemas de monitoreo.

## 02 · Mecánica 1 · pregunta evaluativa (tipo + alcance + Sinergia DNP)

La pregunta evaluativa determina todo lo demás. Una evaluación que intenta contestar *¿el programa funcionó?* sin distinguir tipo de pregunta ni alcance temporal termina midiendo lo más fácil, no lo más importante. El módulo te pide cuatro cosas:

- **Tipo de pregunta.** Cinco opciones canónicas: descripción (¿qué está pasando?), atribución causal (¿la política causó el cambio?), valor (¿vale la pena lo que cuesta?), proceso (¿cómo se está implementando?), gestión (¿la organización está aprendiendo?). Cada tipo exige métodos distintos en el paso 4.
- **Alcance temporal.** Ex-ante (antes de implementar), concurrente (durante), ex-post (después) o meta-evaluación (evaluar evaluaciones previas).
- **Tipología Sinergia DNP.** Seis tipos canónicos del sistema colombiano: *ejecutiva* (capacidad institucional), *operaciones* (cómo se implementa), *resultados* (cumple productos y outcomes), *impacto* (atribución causal del efecto), *institucional* (capacidad estatal) y *mapas de evidencia* (síntesis sistemática · Evidence Gap Maps). Alinear la pregunta con la tipología DNP facilita que el plan sea reconocido por Sinergia.
- **Pregunta principal + sub-preguntas.** Una sola pregunta acotada al grupo afectado, la magnitud esperada del efecto, el horizonte temporal y, si es causal, el contrafáctico. Hasta 5 sub-preguntas auxiliares.

*Las clasificaciones (tipo + alcance) condicionan la sugerencia de método del paso 4. Si cambias el tipo más adelante, la sugerencia se recalcula automáticamente.*

## 03 · Mecánica 2 · teoría de cambio (marco lógico CEPAL)

La teoría de cambio es el puente entre lo que la política *hace* y lo que la política *logra*. Sin ese puente explícito, no hay forma de saber si un resultado se debió a la intervención o al contexto. El módulo usa el marco lógico clásico de CEPAL/ILPES (Ortegón, Pacheco & Prieto, 2005) con cinco niveles:

- **Insumos.** Recursos disponibles: presupuesto, personal, infraestructura, sistemas de información.
- **Actividades.** Lo que hace la política: talleres, capacitaciones, transferencias, regulación.
- **Productos.** Output directo de las actividades: estudiantes con subsidio, capacitaciones dictadas, normas expedidas.
- **Resultados.** Cambios de corto-medio plazo en la población beneficiaria: aumento de asistencia, reducción de deserción.
- **Impacto.** Efecto último sobre el problema. Atribución parcial — el contexto siempre influye.

Adicionalmente se registran **supuestos transversales** (hasta 6): condiciones que deben mantenerse para que la cadena funcione (Mayne, *Contribution Analysis*, 2008+).

*El módulo trae una plantilla seed de educación. Si tu sector es otro, el copiloto IA puede validar tu teoría con plan Premium.*

## 04 · Mecánica 3 · indicadores SMART

Un indicador no es una métrica. Una métrica es cualquier cosa que puedes medir; un indicador es una métrica que captura algo específico de la teoría de cambio. **SMART** = Specific, Measurable, Achievable, Relevant, Time-bound. Si falta uno, el módulo lo marca automáticamente con un chip ámbar que dice cuál letra falta.

Cada indicador tiene 8 campos: nivel (atado a la teoría de cambio), nombre, definición operativa, fórmula explícita, fuente verificable, línea base, meta, frecuencia.

*El copiloto IA puede sugerir 4-6 indicadores SMART derivados de tu pregunta evaluativa y teoría de cambio (plan Pro). Botón + Agregar los inyecta directo al state.*

## 05 · Mecánica 4 · selector de método (14 métodos · frontera 2020-2026)

El método se elige por la pregunta, no por la moda. La versión 2 del módulo trae **catorce métodos** pre-cargados, actualizados a la literatura econométrica reciente (2020-2026). Los *estado del arte* llevan badge especial para distinguirlos de los clásicos:

Método	Autor faro	Pregunta típica
RCT	Banerjee-Duflo-Kremer · JPAL	causal
DID escalonado ★	Callaway-Sant'Anna 2021 · Sun-Abraham 2021	causal
DID clásico (2 períodos)	Card-Krueger 1994	causal
Synthetic Control aumentado ★	Ben-Michael-Feller-Rothstein 2021	causal
Control sintético clásico	Abadie-Diamond-Hainmueller 2010	causal
RDD moderno ★	Cattaneo-Keele-Titiunik 2023	causal
RD clásico	Thistlethwaite-Campbell · Imbens-Lemieux	causal
Double ML ★	Chernozhukov et al. 2018	causal observacional
Causal Forests ★	Wager-Athey 2018 · Athey-Tibshirani 2019	heterogeneidad
Matching / PSM	Rosenbaum-Rubin 1983	causal
Análisis de Contribución ★	Mayne 2024 · WB IEG 2023	valor · proceso · gestión
Cualitativo	Patton 2022 · Yin 2018	valor · proceso
Mixto (QUANT + QUAL)	Creswell-Plano Clark 2017	todo tipo
Value-for-Money + MVPF	HM Treasury 2022 · Hendren NBER 2020	valor · gestión

★ = *método estado del arte (2018-2024)*. La diferencia central frente a v1: el módulo ahora detecta automáticamente cuando el tratamiento es **escalonado** (varias unidades entrando al programa en momentos distintos) y advierte sobre el sesgo TWFE del DID clásico (Goodman-Bacon 2021). En ese caso sugiere el estimador  $ATT(g,t)$  de Callaway-Sant'Anna. El método sugerido aparece con badge salmón según el tipo de pregunta del paso 1. Puedes cambiar la selección — lo importante es que la decisión sea consciente y quede justificada en el plan final.

## 06 · Mecánica 5 · criterios OCDE-DAC

Los seis criterios canónicos (relevance, coherence, effectiveness, efficiency, impact, sustainability) son el lenguaje franco de la evaluación internacional desde 1991, actualizados por la OCDE en 2019. Una evaluación que se posiciona explícitamente frente a cada uno es leída sin fricción por organismos multilaterales y comités técnicos.

El módulo presenta los 6 criterios como cards con su definición OCDE-DAC y un campo de auto-evaluación. Si un criterio no aplica a tu caso, lo puedes marcar con justificación (también es una respuesta válida).

## 07 · Mecánica 6 · análisis económico (CBA · MVPF · CEA)

Tres calculadoras económicas conviven en un mismo paso opcional. La decisión de cuál usar depende de la pregunta y del público lector:

- **CBA · Costo-Beneficio (Green Book HM Treasury 2022).**  $VPN = \sum (B - C) / (1 + r)^t$ . Tasa de descuento configurable (DNP: 9%; Green Book: 3.5%). Horizonte 1-50 años. Ratio B/C como métrica auxiliar. La calculadora reporta los tres resultados (VPN, ratio, NPV/cost) en moneda colombiana formateada.
- **MVPF · Marginal Value of Public Funds (Hendren-Sprung-Keyser 2020).** Beneficios netos para receptores / costo neto al gobierno.  $MVPF > 1$  = política Pareto-superior. Diseñado para comparar programas heterogéneos (transferencias, becas, subsidios) en un solo número, evitando la falsa precisión de monetizar todos los beneficios.
- **CEA · Cost-Effectiveness (J-PAL).** Costo total / outcome total expresado en unidad natural (ej. años adicionales de escolaridad, vidas salvadas, casos prevenidos). Útil cuando monetizar el beneficio es éticamente controvertido o técnicamente imposible.

*El paso es opcional (toggle activo/inactivo). Si se activa, los tres cálculos quedan incluidos en el plan .md final y en el Pre-Analysis Plan exportable.*

## 08 · Mecánica 7 · plan operativo + 3 descargas

El módulo cierra con cuatro campos finales: cronograma estimado, equipo evaluador (con dedicaciones), presupuesto estimado y plan de uso de los resultados. Estos son los campos que típicamente diferencian un plan leído de uno archivado.

El entregable son tres descargas:

- **Plan de evaluación (.md).** Documento estructurado en 6 secciones con todo el state: pregunta, teoría, indicadores, método, criterios DAC, plan operativo. Compatible con formato Sinergia/DNP.
- **Pre-Analysis Plan (.md).** Documento estilo AEA RCT Registry / OSF, con hipótesis primarias y secundarias, especificación econométrica explícita, corrección por hipótesis múltiples pre-registrada y protocolo de desviaciones. Ver sección 09.
- **Matriz de indicadores (.csv).** Tabla cruda con SMART score y missing letters. Útil para llevar a Excel, validar con el equipo o cargar en un sistema de monitoreo.

## 09 · Pre-Analysis Plan exportable (AEA RCT Registry / OSF)

El Pre-Analysis Plan (PAP) es el blindaje contra el p-hacking. Se registra *antes* de recolectar los datos finales; los hallazgos se reportan contra ese plan, no contra una versión revisada. La versión 2 del módulo exporta un PAP listo para subir a [socialscienceregistry.org](https://www.socialscienceregistry.org) (AEA) o a OSF Registries.

Estructura del PAP generado (13 secciones):

- Research question + tipología Sinergia DNP + tipo + alcance.
- Hipótesis primarias (H1.k por outcome de impacto/resultado) y secundarias (H2.k por outcome de producto/actividad).
- Outcomes pre-registrados: primarios (impacto/resultados) y secundarios (productos/actividades) con línea base, meta y frecuencia.
- Teoría de cambio (CEPAL/ILPES) con supuestos críticos.
- Estrategia de identificación + especificación econométrica explícita ( $Y_{it} = \alpha_i + \lambda_t + \beta \cdot T_{it} + \varepsilon$ ; ATT(g,t) si DID escalonado; pesos ridge si SC aumentado; etc.).
- **Corrección por hipótesis múltiples (MHT)**. Bonferroni con  $k \leq 3$  outcomes; Holm o Romano-Wolf con  $k \leq 8$ ; Benjamini-Hochberg  $FDR \leq 0.10$  con  $k \geq 9$ . Anderson 2008 JASA · List-Shaikh-Xu 2019 Exp Econ.
- Heterogeneidad pre-especificada: sexo, edad, territorio, valor línea base, etnia, dosis-respuesta. Cualquier subgrupo no listado se reporta como exploratorio.
- Cálculo de poder (placeholder con inputs: take-up, ICC, atrición, N tratamiento/control, SD baseline).
- Análisis económico (si el toggle del paso 6 está activo): CBA + MVPF + CEA con análisis de sensibilidad sobre la tasa de descuento.
- Criterios OCDE-DAC (pre-compromiso cualitativo).
- Limitaciones conocidas + **protocolo de desviaciones**: cualquier ajuste post-registro debe documentarse en un addendum fechado antes del unblinding; lo contrario se reporta como exploratorio (Olken 2015 JEP).
- Referencias y anclajes metodológicos (AEA · BITSS · OECD-DAC · DNP Sinergia · Mayne · Callaway-Sant'Anna · Cattaneo · Hendren).

## 10 · Copiloto IA y cómo aprovecharlo bien

El módulo incluye tres acciones del copiloto IA distribuidas en el flow:

- **Sugerir indicadores** (Pro+). En el paso 3, dado tu pregunta y tu teoría de cambio, propone 4-6 indicadores SMART listos para agregar con un click.
- **Revisar teoría de cambio** (Premium+). En el paso 2, detecta saltos lógicos, supuestos implícitos, niveles desbalanceados e impactos vagos. Sugiere supuestos transversales faltantes.
- **Generar lectura del plan** (Premium+). En el paso final, interpreta la coherencia pregunta→método→indicadores→criterios. Identifica fortalezas, riesgos del método y puntos a cerrar antes de comité.

*El copiloto sugiere; el humano decide. Las propuestas vienen con botón + Agregar para inyectarlas al state, pero revisa y edita antes de defender el plan ante un comité — el modelo no conoce tu contexto operativo.*

## 11 · Cómo encadenar con los demás módulos del lab

El módulo de Evaluación es el cierre natural del Lab. Idealmente llega al final: tienes el problema enmarcado, las palancas del sistema identificadas y los actores mapeados. Cuándo encadenar:

- **Antes:** si tu política no tiene un problema bien enmarcado, evaluar es evaluar la respuesta equivocada. Empieza por *problema público*.
- **Paralelo:** tu teoría de cambio se beneficia del *análisis estructural* (qué variables mueven el sistema) — las palancas identificadas son candidatos a indicadores de resultado.
- **Después:** el plan de evaluación debe ser defendido ante actores con intereses. *Mactor* te mapea quiénes son y dónde tienes que negociar el alcance o el método.

El lab está pensado para encadenar. Los cuatro módulos comparten chasis visual y la información del state se pasa manualmente entre uno y otro (no automáticamente — eso es por diseño, para que el re-trabajo del analista sea consciente).