

# Matriz de evaluaciones en ciencias sociales

Fit entre diseño de investigación y propósito evaluativo · OnceOnce · once-once.org

Diseñado para esto	Posible con condiciones	No aplica
--------------------	-------------------------	-----------

	<b>Impacto</b> <i>¿El programa causó el cambio?</i>	<b>Mecanismos</b> <i>¿Por qué y cómo funciona?</i>	<b>Heterogeneidad</b> <i>¿Para quién funciona?</i>	<b>Costo-efectividad</b> <i>¿Cuánto impacto por dólar?</i>	<b>Implementación</b> <i>¿Se ejecutó como fue diseñado?</i>
<b>Experimental (RCT)</b> Aleatorización individual o por cluster. Control directo sobre la asignación.	<b>Diseñado para esto</b> La aleatorización garantiza que las diferencias en resultados se deben al programa. Validez interna máxima.	<b>Posible con condiciones</b> Requiere medir variables mediadoras y aplicar análisis de mediación (Imai et al.). No surge del RCT base.	<b>Posible con condiciones</b> Requiere potencia estadística por subgrupo y corrección por comparaciones múltiples. Idealmente pre-especificado.	<b>Posible con condiciones</b> El RCT establece el efecto; el análisis de costos y la unidad comparable (DALY, QALY) van por separado.	<b>Posible con condiciones</b> Se puede medir fidelidad dentro del RCT, pero requiere instrumentos adicionales. Componente complementario.
<b>Cuasi-experimental</b> DiD, RDD, IV, Control sintético, ITS, PSM. Control estadístico.	<b>Posible con condiciones</b> Válido solo si se cumplen los supuestos del diseño (tendencias paralelas en DiD, continuidad en RDD, exclusión en IV).	<b>Posible con condiciones</b> Las variables mediadoras suelen ser endógenas. Requiere instrumentos adicionales o diseños muy específicos.	<b>Posible con condiciones</b> Factible si el diseño tiene suficiente variación en los subgrupos y se mantienen los supuestos de identificación.	<b>Posible con condiciones</b> El diseño establece el efecto; el análisis de costos va por separado. La incertidumbre en el efecto afecta la estimación.	<i>No aplica</i> No está construido para medir procesos internos de implementación. Eso requiere datos cualitativos o encuestas de proceso.
<b>Observacional</b> Longitudinal, corte transversal, ITS sin control. Sin manipulación.	<i>No aplica</i> Sin control sobre la asignación, no es posible descartar explicaciones alternativas. Asociación, no causalidad.	<i>No aplica</i> Si no se establece el efecto causal base, tampoco se puede identificar el mecanismo que lo produce.	<b>Posible con condiciones</b> Útil para describir variación en resultados por subgrupos, pero sin identificación causal. Genera hipótesis.	<i>No aplica</i> Sin efecto causal establecido, la relación costo-resultado no tiene interpretación válida de costo-efectividad.	<b>Diseñado para esto</b> El más adecuado para documentar cómo se implementa un programa: qué ocurre en la práctica, dónde hay brechas.
<b>Cualitativo causal</b> Process tracing, QCA, Realist evaluation.	<i>No aplica</i> No está diseñado para estimar efectos promedio poblacionales. Su fortaleza es la causalidad dentro del caso.	<b>Diseñado para esto</b> Único diseño que traza la cadena causal paso a paso (process tracing) o identifica configuraciones necesarias y suficientes (QCA).	<b>Diseñado para esto</b> QCA identifica en qué configuraciones de contexto y condiciones se produce un resultado — definición precisa de heterogeneidad causal.	<i>No aplica</i> Sin estimación de efectos poblacionales ni datos de costos sistematizados, no es posible calcular costo-efectividad.	<b>Posible con condiciones</b> El process tracing puede documentar en detalle cómo se implementó un programa y por qué se produjeron desviaciones.
<b>Meta-análisis y síntesis</b> Combina múltiples estudios. Estándares Campbell y Cochrane.	<b>Diseñado para esto</b> Sintetiza estimaciones de efecto de múltiples estudios para producir una estimación más precisa entre contextos.	<b>Posible con condiciones</b> Solo si los estudios primarios midieron las mismas variables mediadoras y reportan los datos necesarios.	<b>Diseñado para esto</b> Puede estimar si los efectos varían por población, contexto o características del programa con mayor poder estadístico.	<b>Diseñado para esto</b> Permite comparar costo-efectividad entre intervenciones distintas con métrica común. Es la base de lo que hace GiveWell.	<i>No aplica</i> La síntesis trabaja con resultados de estudios, no con procesos de implementación de cada uno.

# Matriz de evaluaciones en ciencias sociales — Referencias bibliográficas

Papers representativos por diseño de investigación y propósito evaluativo · OnceOnce · once-once.org

	Paper que usa exactamente este diseño para esta pregunta	Paper que usa este diseño para esta pregunta con análisis adicional	Sin combinación metodológicamente válida		
	<b>Impacto</b> <i>¿El programa causó el cambio?</i>	<b>Mecanismos</b> <i>¿Por qué y cómo funciona?</i>	<b>Heterogeneidad</b> <i>¿Para quién funciona?</i>	<b>Costo-efectividad</b> <i>¿Cuánto impacto por dólar?</i>	<b>Implementación</b> <i>¿Se ejecutó como fue diseñado?</i>
<b>Experimental (RCT)</b> Aleatorización individual o por cluster.	<b>Duflo, Dupas &amp; Kremer (2015)</b> <i>American Economic Review</i> RCT en 328 escuelas de Kenya. Efecto causal de la contratación de docentes locales sobre resultados de aprendizaje.	<b>Imai, Keele &amp; Tingley (2010)</b> <i>Psychological Methods</i> Formalización del análisis de mediación causal aplicado a datos de RCT. Estándar para identificar mecanismos dentro de experimentos.	<b>Athey &amp; Imbens (2016)</b> <i>Proceedings of the National Academy of Sciences</i> Causal forests para identificar subgrupos con efectos heterogéneos en datos de RCT. Fundamento del HTE moderno.	<b>Kremer, Glennerster &amp; Muralidharan (2013)</b> <i>Handbook of the Economics of Education</i> RCT con análisis de costo por año de escolaridad adicional. Marco de referencia para costo-efectividad en educación.	<b>Banerjee et al. (2015)</b> <i>Science — J-PAL</i> RCT en 6 países con medición de fidelidad de implementación integrada. El programa "graduation" de Bandhan Trust.
<b>Cuasi-experimental</b> DiD, RDD, IV, Control sintético, ITS, PSM.	<b>Card &amp; Krueger (1994)</b> <i>American Economic Review</i> DiD clásico: efecto causal del salario mínimo en NJ sobre el empleo, comparando con PA como control.  <b>Thistlethwaite &amp; Campbell (1960)</b> <i>Journal of Educational Psychology</i> Fundacional del RDD: efecto de becas universitarias usando el umbral de elegibilidad como discontinuidad.	<b>Angrist &amp; Krueger (1991)</b> <i>Quarterly Journal of Economics</i> Variables instrumentales: fecha de nacimiento como instrumento para escolaridad. Mecanismo de retorno a la educación.	<b>Imbens &amp; Lemieux (2008)</b> <i>Journal of Econometrics</i> Guía práctica del RDD para estimar efectos heterogéneos en torno al umbral de elegibilidad de los programas.	<b>Dhalwal et al. (2013)</b> <i>Abdul Latif Jameel Poverty Action Lab</i> DiD + análisis de costos: comparación de costo por resultado en intervenciones educativas de distintos países.	<i>El diseño cuasi-experimental mide resultados, no procesos internos.</i>
<b>Observacional</b> Longitudinal, corte transversal, ITS sin control.	<i>Sin control sobre la asignación no es posible establecer causalidad.</i>	<i>Sin efecto causal base no puede identificarse el mecanismo que lo produce.</i>	<b>Deaton &amp; Cartwright (2018)</b> <i>Social Science &amp; Medicine</i> Análisis observacional de heterogeneidad en efectos de RCTs: por qué los efectos promedio ocultan variación relevante por subgrupo.	<i>Sin efecto causal establecido la relación costo-resultado no es interpretable como costo-efectividad.</i>	<b>Fixsen et al. (2005)</b> <i>University of South Florida — NIRN</i> Revisión longitudinal de 25 años de literatura sobre fidelidad de implementación. Marco fundacional de implementation science.
<b>Cualitativo causal</b> Process tracing, QCA, Realist evaluation.	<i>No está diseñado para estimar efectos promedio poblacionales.</i>	<b>Beach &amp; Pedersen (2013)</b> <i>University of Michigan Press</i> Process-tracing: inferencia causal cualitativa que traza la cadena de evidencia dentro del caso para testear mecanismos.  <b>Ragin (2008)</b> <i>University of Chicago Press</i> Redesigning Social Inquiry: QCA como método para identificar configuraciones de condiciones necesarias y suficientes.	<b>Schneider &amp; Wagemann (2012)</b> <i>Cambridge University Press</i> Set-theoretic methods: QCA para identificar en qué configuraciones de contexto se produce un resultado — heterogeneidad causal.	<i>Sin estimación de efectos poblacionales ni costos sistematizados no es posible calcular costo-efectividad.</i>	<b>Pawson &amp; Tilley (1997)</b> <i>Sage Publications</i> Realistic evaluation: mecanismo + contexto + resultado. Process tracing aplicado a la implementación de programas sociales.

	<b>Impacto</b> <i>¿El programa causó el cambio?</i>	<b>Mecanismos</b> <i>¿Por qué y cómo funciona?</i>	<b>Heterogeneidad</b> <i>¿Para quién funciona?</i>	<b>Costo-efectividad</b> <i>¿Cuánto impacto por dólar?</i>	<b>Implementación</b> <i>¿Se ejecutó como fue diseñado?</i>
<b>Meta-análisis y síntesis</b> Combina múltiples estudios. Estándares Campbell y Cochrane.	<b>Gechter &amp; Meager (2024)</b> <i>NBER Working Paper</i> Meta-análisis bayesiano de RCTs: cómo generalizar estimaciones de impacto entre sitios con heterogeneidad en los efectos.	<b>Vivalt (2020)</b> <i>Journal of the European Economic Association</i> 'How much can we generalize from impact evaluations?' Meta-análisis de 600+ estudios: heterogeneidad de efectos entre contextos.	<b>Vivalt (2020)</b> <i>Journal of the European Economic Association</i> El mismo análisis permite estimar si los efectos varían sistemáticamente por tipo de población, región o características del programa.	<b>GiveWell Cost-Effectiveness Analysis (2023)</b> <i>GiveWell — Annual Report</i> Meta-análisis de evidencia de RCTs para comparar costo por DALY evitado entre intervenciones de salud global. Estándar del campo.	<i>La síntesis trabaja con resultados de estudios, no con procesos de implementación de cada uno.</i>