

Typology of Evaluation Approaches

Three perspectives: method, cycle, instrument

Reference note for the design of Themiros core — September 2026

Classification Framework

Evaluation approaches are classified along three largely independent dimensions. Any real evaluation combines all three.

Dimension	Question	Main Categories
Methodological	How are findings and causality established?	Experimental, quasi-experimental, theory-based, participatory, economic, mixed
Policy or project cycle	When is the evaluation conducted?	Ex-ante, evaluability, formative, mid-term, final, ex-post
Instrument	What is being evaluated?	Project, programme, policy, strategy, sector, theme, portfolio, institution

Example of a combination: an *ex-post* evaluation (cycle) of the impact of a *programme* (instrument) using a *quasi-experimental* design (method).

Two cross-cutting distinctions complete the framework: **who conducts the evaluation**, and **according to which criteria it judges the intervention**. They are addressed at the end of the document.

1. Methodological Perspective: How to Establish Causality and Value

The main dividing line opposes **attribution** and **contribution**. Attribution asks whether the intervention caused the change, and to what extent. Contribution asks how and why the

intervention plausibly contributed to it, among other factors (Mayne 2001, 2012; Befani & Mayne 2014).

The choice of design depends on the attributes of the intervention and the available data, according to the matrix proposed by Stern et al. (2012). The distinction between *complicated* and *complex* aspects structures this entire section (Rogers 2008).

1.1 Attribution Approaches (Counterfactual)

Family	Principle	Strengths	Limitations
Experimental	Random assignment of units to treatment and control groups. Any difference in outcomes is attributable to the intervention.	Maximum internal validity.	High cost. Ethical and political constraints. Poorly suited to national policies and complex interventions.
Quasi-experimental	Construction of a credible counterfactual without randomization.	Applicable <i>ex post</i> and at the scale of a territory.	Depends on data quality and identification assumptions.

Method	Family	Principle	Reference
Randomized controlled trial (RCT)	Experimental	Individual randomization. Variants: cluster RCTs, multi-arm or factorial trials.	Duflo, Glennerster & Kremer 2007
Difference-in-differences	Quasi-experimental	Compares changes over time between treated and untreated groups.	Bertrand, Duflo & Mullainathan 2004
Regression discontinuity	Quasi-experimental	Exploits an eligibility threshold, such as a poverty score.	Imbens & Lemieux 2008
Propensity score matching	Quasi-experimental	Matches each beneficiary with a statistically similar non-beneficiary.	Rosenbaum & Rubin 1983
Synthetic control	Quasi-experimental	Constructs a weighted "twin" from comparison units. Useful for regional or national policies.	Abadie, Diamond & Hainmueller 2010

Method	Family	Principle	Reference
Instrumental variables	Quasi-experimental	Isolates an exogenous source of variation in exposure to the intervention.	Angrist, Imbens & Rubin 1996
Interrupted time series	Quasi-experimental	Detects a break in trend at the time of the intervention.	Bernal, Cummins & Gasparrini 2017

Overview of data requirements and use cases: **Gertler et al. (2016)**.

References — §1.1

Abadie, A., Diamond, A. & Hainmueller, J. (2010). Synthetic Control Methods for Comparative Case Studies. *JASA*, 105(490), 493-505. doi:10.1198/jasa.2009.ap08746

Angrist, J. D., Imbens, G. W. & Rubin, D. B. (1996). Identification of Causal Effects Using Instrumental Variables. *JASA*, 91(434), 444-455. doi:10.1080/01621459.1996.10476902

Bernal, J. L., Cummins, S. & Gasparrini, A. (2017). Interrupted time series regression... *Int. J. Epidemiology*, 46(1), 348-355. doi:10.1093/ije/dyw098

Bertrand, M., Duflo, E. & Mullainathan, S. (2004). How Much Should We Trust Differences-in-Differences Estimates? *QJE*, 119(1), 249-275. doi:10.1162/003355304772839588

Duflo, E., Glennerster, R. & Kremer, M. (2007). Using Randomization in Development Economics Research: A Toolkit. *NBER TW 333*. nber.org/papers/t0333

Gertler, P. J. et al. (2016). *Impact Evaluation in Practice*, 2nd ed. World Bank & IDB. doi:10.1596/978-1-4648-0779-4

Imbens, G. W. & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *J. Econometrics*, 142(2), 615-635. doi:10.1016/j.jeconom.2007.05.001

Rosenbaum, P. R. & Rubin, D. B. (1983). The central role of the propensity score... *Biometrika*, 70(1), 41-55. doi:10.1093/biomet/70.1.41

1.2 Contribution Approaches (Theory-Based)

These approaches open the "black box" by testing the causal chain described in a theory of change (**Weiss 1997; White 2009**). They are suitable for complex interventions, small numbers of cases, and policies where no counterfactual is possible (**Rogers 2008; Stern et al. 2012**).

Approach	Origin	Central Question	Reference
Contribution analysis	Mayne	Did the intervention plausibly contribute to the observed change?	Mayne 2001, 2012
Realist evaluation	Pawson and Tilley	What works, for whom, in what context, and why?	Pawson & Tilley 1997, 2004
Process tracing	Political science	What mechanisms link the intervention to the observed effects?	Collier 2011; World Bank IEG
Qualitative comparative analysis (QCA)	Ragin	What combinations of conditions produce the outcome?	Ragin 2008 (<i>outside MVP</i>)
General elimination method	Scriven	What alternative explanations can be ruled out?	Scriven 2008
Outcome harvesting	Wilson-Grau	What changes occurred and how did the intervention contribute to them? Inductive approach.	Wilson-Grau & Britt 2012/2013
Most significant change	Davies and Dart	What changes do stakeholders consider most important? Inductive approach.	Davies & Dart 2005

The combination of *process tracing* + contribution analysis is executed by the Themiros engine (**Befani & Mayne 2014**), with Bayesian updating for corpus confidence (**Befani & Stedman-Bryce 2017**).

References — §1.2

- Befani, B. & Mayne, J. (2014). Process Tracing and Contribution Analysis. *IDS Bulletin*, 45(6), 17-36. doi:10.1111/1759-5436.12110
- Befani, B. & Stedman-Bryce, G. (2017). Process Tracing and Bayesian Updating for impact evaluation. *Evaluation*, 23(1), 42-60. doi:10.1177/1356389016654584
- Collier, D. (2011). Understanding Process Tracing. *PS: Political Science and Politics*, 44(4), 823-830. doi:10.1017/S1049096511001429
- Davies, R. & Dart, J. (2005). *The 'Most Significant Change' (MSC) Technique: A Guide to Its Use*. man.de.co.uk
- Independent Evaluation Group, World Bank. *Process Tracing Method for Program Evaluation*. ieg.worldbankgroup.org

- Mayne, J. (2001). Addressing attribution through contribution analysis. *Canadian Journal of Program Evaluation*, 16(1), 1-24. doi:10.3138/cjpe.016.001
- Mayne, J. (2012). Contribution analysis: Coming of age? *Evaluation*, 18(3), 270-280. doi:10.1177/1356389012451663
- Pawson, R. & Tilley, N. (1997). *Realistic Evaluation*. Sage. — (2004). *Realist Evaluation*. betterevaluation.org
- Ragin, C. C. (2008). *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. University of Chicago Press. ISBN 978-0-226-70275-9.
- Scriven, M. (2008). A Summative Evaluation of RCT Methodology... *Journal of MultiDisciplinary Evaluation*, 5(9), 11-24. doi:10.56645/jmde.v5i9.160
- Weiss, C. H. (1997). Theory-based evaluation: Past, present, and future. *New Directions for Evaluation*, 76, 41-55. doi:10.1002/ev.1086
- White, H. (2009). Theory-based impact evaluation. *J. Development Effectiveness*, 1(3), 271-284. doi:10.1080/19439340903114628
- Wilson-Grau, R. & Britt, H. (2012/2013). *Outcome Harvesting*. Ford Foundation. usaidlearninglab.org

1.3 Other Families

Family	Approaches	Use	Reference
Participatory and use-oriented	Utilization-focused evaluation (Patton). Developmental evaluation. Empowerment evaluation.	Innovative or adaptive programmes. Ownership of findings by stakeholders.	Patton 2008 (<i>outside MVP</i>)
Economic	Cost-benefit analysis (CBA). Cost-effectiveness analysis (CEA). Cost-utility analysis. Social return on investment (SROI). Multi-criteria analysis.	Relationship between costs and outcomes. Multi-criteria analysis is essential when benefits cannot be monetized, a frequent case in climate adaptation.	OECD 2021; Gertler et al. 2016
Case studies	Single case, multiple cases, comparative cases.	In-depth understanding of a context.	Yin (case studies, <i>outside MVP</i>)
Systemic	Complexity-sensitive approaches. Systems mapping.	Interventions with non-linear effects and multiple actors.	Rogers 2008

Family	Approaches	Use	Reference
Evidence synthesis	Systematic reviews. Meta-analyses. Rapid reviews.	Cross-cutting capitalization from existing evaluations.	—
Mixed methods	Combination of quantitative and qualitative methods.	Current standard: quasi-experimental estimation of the effect, complemented by a theory-based explanation of mechanisms.	White 2009; Stern et al. 2012

References — §1.3

OECD (2021). *Applying Evaluation Criteria Thoughtfully*. OECD Publishing, Paris. doi:10.1787/543e84ed-en

Patton, M. Q. (2008). *Utilization-Focused Evaluation*, 4th ed. Sage.

Rogers, P. J. (2008). Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions. *Evaluation*, 14(1), 29-48. doi:10.1177/1356389007084674

Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R. & Befani, B. (2012). *Broadening the range of designs and methods for impact evaluations*. DFID Working Paper 38. gov.uk

2. Cycle Perspective: When to Evaluate

Each stage of the policy or project cycle calls for a specific type of evaluation and purpose.

Stage	Type of Evaluation	Main Purpose
Design	Ex-ante evaluation. Feasibility study (<i>appraisal</i>). Impact study. Regulatory impact assessment (RIA).	Justify and design the intervention.
Before launch	Evaluability assessment. Baseline study.	Verify that the intervention is evaluable. Establish reference values.
Implementation	Monitoring. Formative or process evaluation. Real-time evaluation (common in humanitarian contexts).	Monitor progress. Improve implementation.

Stage	Type of Evaluation	Main Purpose
Mid-term	Mid-term review or evaluation.	Reorient. Reallocate resources.
Closure	Final evaluation. Completion report (e.g., the World Bank's ICR).	Assess results against objectives.
After closure	Ex-post evaluation. Impact evaluation. Sustainability evaluation.	Measure lasting effects. Accountability. Learning.
Cross-cutting	Meta-evaluation. Synthesis evaluation.	Assess the quality of evaluations. Aggregate lessons learned.

Distinction	Content
Formative / summative	Formative improves the intervention during implementation. Summative judges its overall value, generally at the end.
"Evaluate first" (Better Regulation, EU)	Ex-post evaluation of existing rules precedes any new ex-ante impact assessment.

References — §2

OECD (2021). *Applying Evaluation Criteria Thoughtfully*. doi:10.1787/543e84ed-en

European Commission. *Better Regulation Guidelines*. commission.europa.eu

3. Instrument Perspective: What to Evaluate

The object evaluated gradually expands from the project to the policy. The broader it becomes, the less feasible randomization is (**Stern et al. 2012**).

Object	Scope	Methodological Remark
Project	Single intervention, bounded in time, with defined outputs.	Prime ground for counterfactual methods.
Programme	Set of linked projects sharing common objectives.	Often combines several methods.

Object	Scope	Methodological Remark
Public policy	Policy, law, or regulation, often at national scale.	Counterfactual difficult. Theory-based approaches and synthetic control are common.
Strategy	Strategic framework of an organization or government.	Coherence and positioning analysis.
Sector	Entire sector (agriculture, health), including sector-wide approaches.	Multiplicity of actors and instruments.
Theme	Cross-cutting theme (gender, climate adaptation) across multiple interventions.	Synthesis approach.
Country or regional programme	Entire support of a donor to a country or region.	Relevance and coherence at the forefront.
Portfolio	Group of investments or operations.	Aggregate performance analysis.
Institution or organization	Performance, relevance, and governance of an organization.	Organizational criteria.
Specific instrument	Aid modality or policy instrument: budget support (OECD DAC three-step approach), results-based financing, subsidies, taxes.	Methodological framework specific to each instrument.
Fitness check (EU)	Group of related policies.	Coherence, overlaps, gaps.

References — §3

Stern, E. et al. (2012). *Broadening the range of designs and methods for impact evaluations*. DFID WP 38. [gov.uk](https://www.gov.uk)

4. Cross-Cutting Distinctions

Classification Criterion	Categories
Who evaluates	Internal evaluation. External or independent evaluation. Joint evaluation (multiple donors or partners). Self-evaluation. Peer review.
Criteria applied	OECD DAC criteria: relevance, coherence (added in 2019), effectiveness, efficiency, impact, sustainability.

The definitions of the six criteria and their thoughtful application are specified by **OECD (2021)**.

References — §4

OECD (2021). *Applying Evaluation Criteria Thoughtfully*. OECD Publishing, Paris. doi:10.1787/543e84ed-en

5. Articulation of the Three Dimensions

The choice of approach depends on four factors.

Factor	Implication
The evaluation question	"How much?" calls for an experimental or quasi-experimental design. "How and why?" calls for a theory-based approach.
The nature of the intervention	Simple projects and complex or adaptive policies require different methods (Rogers 2008).
Data availability	A usable baseline and a comparison group broaden the options.
Budget, timeline, and intended use	These constraints bound the feasible design.

As the object moves from project to programme, then to policy, randomization becomes less feasible. Theory-based approaches, mixed methods, and syntheses then become the norm (**Stern et al. 2012; White 2009**).

General Bibliography

Theoretical and methodological foundation

- Befani, B. & Mayne, J. (2014). Process Tracing and Contribution Analysis. *IDS Bulletin*, 45(6), 17-36. doi:10.1111/1759-5436.12110
- Befani, B. & Stedman-Bryce, G. (2017). Process Tracing and Bayesian Updating for impact evaluation. *Evaluation*, 23(1), 42-60. doi:10.1177/1356389016654584
- Collier, D. (2011). Understanding Process Tracing. *PS: Political Science and Politics*, 44(4), 823-830. doi:10.1017/S1049096511001429
- Davies, R. & Dart, J. (2005). *The 'Most Significant Change' (MSC) Technique: A Guide to Its Use*. man de.co.uk
- Mayne, J. (2001). Addressing attribution through contribution analysis. *Canadian Journal of Program Evaluation*, 16(1), 1-24. doi:10.3138/cjpe.016.001
- Mayne, J. (2012). Contribution analysis: Coming of age? *Evaluation*, 18(3), 270-280. doi:10.1177/1356389012451663
- OECD (2021). *Applying Evaluation Criteria Thoughtfully*. OECD Publishing, Paris. doi:10.1787/543e84ed-en
- Pawson, R. & Tilley, N. (1997). *Realistic Evaluation*. Sage. — (2004). *Realist Evaluation*. betterevaluation.org
- Ragin, C. C. (2008). *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. University of Chicago Press. ISBN 978-0-226-70275-9.
- Rogers, P. J. (2008). Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions. *Evaluation*, 14(1), 29-48. doi:10.1177/1356389007084674
- Scriven, M. (2008). A Summative Evaluation of RCT Methodology... *Journal of MultiDisciplinary Evaluation*, 5(9), 11-24. doi:10.56645/jmde.v5i9.160
- Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R. & Befani, B. (2012). *Broadening the range of designs and methods for impact evaluations*. DFID Working Paper 38. gov.uk
- Weiss, C. H. (1997). Theory-based evaluation: Past, present, and future. *New Directions for Evaluation*, 76, 41-55. doi:10.1002/ev.1086
- White, H. (2009). Theory-based impact evaluation. *J. Development Effectiveness*, 1(3), 271-284. doi:10.1080/19439340903114628
- Wilson-Grau, R. & Britt, H. (2012/2013). *Outcome Harvesting*. Ford Foundation. usaidlearninglab.org

Counterfactual methods

- Abadie, A., Diamond, A. & Hainmueller, J. (2010). Synthetic Control Methods for Comparative Case Studies. *JASA*, 105(490), 493-505. doi:10.1198/jasa.2009.ap08746
- Angrist, J. D., Imbens, G. W. & Rubin, D. B. (1996). Identification of Causal Effects Using Instrumental Variables. *JASA*, 91(434), 444-455. doi:10.1080/01621459.1996.10476902

Bernal, J. L., Cummins, S. & Gasparrini, A. (2017). Interrupted time series regression... *Int. J. Epidemiology*, 46(1), 348-355. doi:10.1093/ije/dyw098

Bertrand, M., Duflo, E. & Mullainathan, S. (2004). How Much Should We Trust Differences-in-Differences Estimates? *QJE*, 119(1), 249-275. doi:10.1162/003355304772839588

Duflo, E., Glennerster, R. & Kremer, M. (2007). Using Randomization in Development Economics Research: A Toolkit. *NBER TW 333*. nber.org/papers/t0333

Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B. & Vermeersch, C. M. J. (2016). *Impact Evaluation in Practice*, 2nd ed. World Bank & IDB. doi:10.1596/978-1-4648-0779-4

Imbens, G. W. & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *J. Econometrics*, 142(2), 615-635. doi:10.1016/j.jeconom.2007.05.001

Rosenbaum, P. R. & Rubin, D. B. (1983). The central role of the propensity score... *Biometrika*, 70(1), 41-55. doi:10.1093/biomet/70.1.41

Outside MVP

Independent Evaluation Group, World Bank. *Process Tracing Method for Program Evaluation*. ieg.worldbankgroup.org

Patton, M. Q. (2008). *Utilization-Focused Evaluation*, 4th ed. Sage.