International Initiative for Impact Evaluation

Annex 2: Evaluation tools that can be used in disaggregated component evaluation in Step 2 and 3 of the 5-step complexity-responsive evaluation framework

1. Theory-driven approaches	
Program theory models	Articulates how a program is intended to operate and how outcomes will be achieved. Planned processes and outcomes are defined as the counterfactual and compared with monitored outcomes.
Historical analysis	Economic and political historians address 'what if' questions by comparing historical events with plausible alternatives.
General elimination theory	The crime investigation approach is used to identify a list of possible causes of an outcome. These are assessed and implausible causes are eliminated.
Contribution analysis	Uses program theory to develop and test a plausible story as to how the program may have contributed to the observed outcomes. This is tested through a 7-step approach that identifies and tests alternative explanations (rival hypotheses) of why the observed changes have occurred. (Mayne 2012; Leeuw 2012)
2. Quantitative approaches	
Experimental designs	Subjects are randomly assigned to treatment (project) and comparison groups that are compared before and after the treatment. This is the only design that uses an unbiased statistical counterfactual.
Quasi-experimental designs [QED]	In the majority of development programs random assignment is not possible and the best approximation is the use of matched comparison groups. In the strongest designs, groups are statistically matched using techniques such as propensity score matching. Weaker designs use judgmental matching.
Pipeline designs and natural experiments	Widely used versions of QEDs. Pipeline designs use the fact that many projects are implemented in clearly defined phases (e.g. road construction or water supply projects) to use subjects who will enter the project in phase 2 as a comparison for subjects entering in phase 1. While pipeline designs use planned phasing, natural experiments take advantage of unplanned delays, or of similar projects implemented by other agencies, for the comparison.
Concept mapping	Stakeholders or experts help define indicators of project outcomes. These are developed into scales that are then used (by stakeholders or experts) to rate the performance of different projects, programs or country policies. The approach can be used for pre-test post-test

	comparisons or it can be used retrospectively to rate the changes in outcome variables over the life of the project
Statistical comparisons of similar countries	Statistical comparisons are made between countries in which programs or policies are being implemented with other similar countries.
Citizen report cards	Surveys are conducted, usually in a particular city, in which residents are asked to rate their experience with public service agencies (water, transport, police etc). Surveys are repeated several years later after agencies have had time to address problems addressed in the first survey.
Social network analysis	SNA analyzes communication networks to compute indicators on the volume, content and structure of communications among agencies involved in a program. The analysis is repeated over time and changes can be calculated in terms of the volume or structure of communications.
Big data science approaches	Multiple sources of big data are becoming available that can be combined to strengthen comparisons between project and comparison groups, and also to assess the influence of external contextual factors. An important development is that longitudinal data sets are becoming available (e.g. satellite and remote sensor images and social media streams such as Twitter).
3. Qualitative and participatory approaches	
Realist evaluation	The approach addresses questions such as 'What works?', 'For whom?', 'When and where?', and 'Why?'. It also focuses on understanding how the program actually works. The approach focuses on context and mechanisms that influence operation and outcomes.
Qualitative comparisons of other countries, regions and sectors	While it is sometimes possible to conduct quantitative comparisons with other countries or regions, in many cases it is only possible to make descriptive comparisons.
PRA and other participatory group methods	The opinions and perceptions of community and other groups are elicited through a range of participatory consultation mechanisms covering, perceived changes over the life of a project (or longer period), the reasons for the changes and which groups have benefited and which have been negatively affected.
Focus groups	Small groups representing different sectors of the target population are interviewed on their experiences and opinions about a project or other intervention. Groups can include both beneficiaries and non-beneficiaries.

Expert judgment and key informants	Representative samples of experts or key informants are interviewed about a project or policy intervention.
Public expenditure tracking	Tracing the steps through which approved public expenditures reach the front-line education, health or other agencies. Estimating the proportion of funds that are not received and which communities are most and least affected.
4. Case studies	
Descriptive case studies	Cases are compared using qualitative and descriptive methods. Cases can use 'thick description' to provide in-depth understanding of lived experiences.
Qualitative comparative analysis (QCA)	A set of binary attributes are compiled for each case. An analysis is conducted to identify the necessary and sufficient configuration of attributes for an outcome to be present or absent.
5. Big data tools and analysis [See Annex 4]	
Source: Adapted from: Bamberger, Vaessen and Raimondo (2016) Dealing with complexity in development evaluation; and, Bamberger and Mabry (2020) Real World Evaluation Chapter 16 Table 16.4	