



Monitoring and evaluation of interventions to prevent and tackle labour exploitation and forced labour in agriculture

Technical note

“Support to strengthening the inter-institutional governance in the field of labour exploitation in Italy”



International
Labour
Organization



Co-funded by
the European Union



Contents

<i>Introduction</i>	<i>1</i>
<i>1. Main features of monitoring systems</i>	<i>2</i>
<i>1.1. Monitoring matrix: Objectives, results and indicators</i>	<i>2</i>
<i>2. Analysis of monitoring data.....</i>	<i>5</i>
<i>3. Ex-post evaluation.....</i>	<i>6</i>
<i>4. Counterfactual evaluation of the effects of public policies</i>	<i>7</i>
<i>4.1. Counterfactual evaluation: experimental designs</i>	<i>7</i>
<i>4.2. Counterfactual evaluation: non-experimental (or quasi-experimental) designs</i>	<i>9</i>
<i>5. Theory-based evaluation.....</i>	<i>11</i>
<i>5.1. The realistic approach to evaluating effects</i>	<i>11</i>
<i>Bibliographic references</i>	<i>15</i>

Introduction

The aim of this technical brief is to provide an overview of the monitoring systems and key methods for the *ex-post* evaluation of the impact of public policy interventions and to offer areas of discussion to the institutions tasked with implementing measures to combat labour exploitation and protect and assist the victims.

Monitoring and evaluation are the processes that allow policymakers to assess how an intervention evolves over time (monitoring); how effectively an intervention was implemented and whether there are gaps between the planned and achieved results (evaluation); and whether the observed changes are due to the intervention alone or to other factors (impact evaluation).

Monitoring is a continuous process of collecting and analysing information about the implementation of an intervention and comparing actual against planned results in order to judge how well the final objective will be accomplished. It uses the data generated through the implementation of an intervention (e.g., resources employed, characteristics of beneficiaries, end of intervention situation) to make comparisons between the achieved and intended results and to assess the differences (for example, between the final beneficiaries, the type of intervention, the geographical areas).

Evaluation is a process that systematically and objectively assesses all the elements of an intervention (e.g., design, implementation and results achieved) to determine its overall worth. The objective is to provide evidence-based information to identify ways to improve the results over time. Broadly speaking, there are two types of evaluation:

- **Performance evaluations** focus on the quality of the implementation and the short-term and medium-term outcomes achieved by an intervention. Although based on information collected through the monitoring system, they have the much broader purpose of determining whether the observed changes are the result of the interventions or due to other factors.
- **Impact evaluations** look for the difference between the actual results that can be attributed to the intervention and the hypothetical results in a “counterfactual” situation, i.e., in the absence of the intervention. Their aim is to separate the changes caused by the intervention from the changes that nevertheless occur in the target group or geographical area concerned by the intervention, over a given period of time.

Monitoring and evaluation should also collect information on the financial resources involved in the intervention, to judge the benefits of an intervention against its costs and identify which intervention has the highest rate of return. Two tools are commonly used:

- A **cost-benefit analysis** estimates the total benefit of an intervention compared to its total costs. This type of analysis is normally used *ex-ante*, to decide among different intervention options. The main difficulty is to assign a monetary value to so-called “intangible” benefits. For example, if the main benefit of a reintegration intervention is the increase of employment and earning opportunities for victims of labour exploitation, these are considered tangible benefits to which a monetary value can be assigned. However, having a job and decent salary also increase the victims’ self-esteem and self-determination, which are more difficult to express in monetary terms since each individual values them differently.

- A **cost-effectiveness analysis** compares the costs of two or more interventions in yielding the same outcome. Take for example the launching of a helpline for the victims of worker exploitation and carrying out inspections on farms; each has the objective to identify the potential victims of exploitation, but the helpline does so at the cost of 10,000 euro per individual, while the inspections cost an estimated 6,000 euro per victim. Inspections therefore perform better – are more cost-effective – than the helpline.

1. Main features of monitoring systems

Monitoring and evaluation are key tools for collecting data and generating reliable information on the performance of interventions and on the institutional actors involved in preventing and combating labour exploitation and supporting the victims. Monitoring highlights the strengths and weaknesses of the corrective measures put into place and the necessary improvements.

A reliable monitoring and evaluation system can also support the policymaking and funding authorities, at the central and local levels, engaged in verifying compliance. A monitoring system can also roll out benchmarks for stakeholders, in order to: (i) track developments over time; (ii) identify possible critical issues that need to be addressed; (iii) allow geographical comparisons; and (iv) measure service quality and user satisfaction.

A prerequisite for ensuring the success of a monitoring system is the involvement of all the actors in defining the key elements of the system and feeding in the information to acquire data at the source. One such key element is the active involvement of the institutional partners tasked with implementing the interventions, as a means for satisfying information needs while ensuring uniformity of content.

1.1. *Monitoring matrix: Objectives, results and indicators*

The monitoring matrix outlines, for each objective, the principal goal to be achieved and related method of measurement, the partner institution responsible for coordinating the interventions, and the underlying rationale.

Performance indicators are concise quantitative and qualitative measures that can be easily determined and are used to track the progress of interventions on a regular basis. The indicators are determined on the basis of the information to be collected and analysed. Quantitative indicators measure changes in a specific value (number, mean or median) and a percentage. Qualitative indicators provide insights into changes in attitudes, beliefs, motives and behaviours of individuals. Although important, information on qualitative indicators is more time-consuming to collect, measure and analyse, especially in the early stages of an intervention implementation.

It should be emphasised that the success of a monitoring and evaluation system depends not just on the amount of data and information generated but also on how the information is used. In this regard, performance indicators in the monitoring process should be SMART, i.e.:

S - *Specific* (measuring only what needs to be measured);

M - *Measurable* (easily collecting the necessary qualitative and quantitative information),

A- *Available* (usable and useful for managers at an acceptable cost);

R- *Relevant* (inherent in the result and able to measure changes over time); and

T- *Timely* (limited in time).

The monitoring matrix features three types of indicators:

1. **Baseline indicators** set the baseline – or current – situation before an intervention is implemented and are based on either secondary (already available) data or the collection of primary data.
2. **Output indicators** provide an ongoing measurement of the relationship between the inputs, activities and outputs of an intervention or measure. Their purpose, in particular, is to make sure that the inputs produce the planned activities and that the activities then produce the planned outputs.
3. **Outcome indicators** measure the short-term results achieved by an intervention through its outputs. They are tracked at specific periods in time (e.g., on an annual basis) to provide a snapshot of the results obtained at that moment. These indicators are most relevant when monitoring public policy interventions, such as interventions against labour exploitation. When an outcome indicator specifies the value (number, mean, median or percentage) that it should reach, the indicator becomes a target. The function of monitoring, therefore, is to regularly check the progress of the intervention, to make sure that it is on track to achieve the planned outcome.

Output and outcome indicators almost always provide for a breakdown of the collected data, indicated directly in the matrix. The most common breakdowns are the characteristics of the intervention beneficiaries (gender, age, nationality); the geographical area; and the type of service provided (Box 1).

Objective to be achieved

Box 1: Example of a monitoring matrix

Action 5: Planning and implementation of decent housing solutions for agricultural workers as an alternative to spontaneous settlements and other degrading accommodation				
Main outcome	Indicator	Definition and measurement	Baselines, targets and data sources	Responsible unit and data collection frequency
5.1. Definition of quality criteria of housing solutions for the design, management and delivery of services at local level Definition of the main outcome	70 % of settled agricultural workers have access to accommodation in line with nationally defined minimum quality standards Value of outcome indicator to be achieved within the period Type of indicator (implementation or outcome) Level of disaggregation of data on the indicator	Number of settled agricultural workers who have access to accommodation in line with minimum quality standards out of the total number of settled workers Definition of the outcome indicator Type of indicator Outcome indicator Unbundling - Characteristics of workers (gender, age, nationality) - Type of accommodation - Geographical area The source of the information needed to measure the value of the indicator	Baseline Information on the number of agricultural workers in degrading housing conditions not aggregated at national level Value of the initial indicator (baseline) on which the measurement is based Target By 2022, 70% of settled agricultural workers have access to housing that meets minimum standards set at national level. Data sources Data from the information system managed by the DG Information Systems of the MLPS) Value of the indicator (target) to be reached within the specified time period	Responsible unit National Association of Italian Municipalities (ANCI) The institutional partner responsible for the collection and analysis of monitoring data Frequency: Annual Frequency of data collection for monitoring

Another group of indicators that can be used when analysing information generated by the monitoring system are context indicators, which are an indirect tool for monitoring the effects of policy interventions because they describe the context in which the interventions are implemented and how it evolves over time.

Besides the indicators, a monitoring matrix includes the units that are responsible for the data collection and analysis, data collection frequency, and data sources. Data collection frequency for monitoring indicators is almost always annual.

2. Analysis of monitoring data

Monitoring data can be analysed in two ways, according to different procedures and timeframes.

One type of analysis, carried out annually, consists of an intervention management tool based on the value of the monitoring matrix indicators. Its aim is to verify the progress of an intervention by assessing what works properly and what doesn't, to provide timely implementation information and data. This analysis is then condensed in a summary report accompanied by a monitoring data matrix.

The second type of analysis is a performance evaluation that focuses on implementation quality and intervention outcomes. This type of evaluation is based on short-term and medium-term results (e.g., the number of labour exploitation victims subsequently reintegrated at work), as well as on the relevance (are the objectives well-suited to solving the relevant problems?), effectiveness (will the results contribute to the achievement of the aim?), and efficiency (will it be possible to achieve the expected results at a reasonable cost?) of the interventions. This evaluation includes analyses and interpretations aimed at expressing a value judgement of the intervention, comparing the results obtained with the intended results, the time taken to achieve them, the costs incurred and the quality of the services. The baseline for this evaluation is the information collected by the monitoring system, with the aim of verifying whether the observed changes are the result of the interventions, as implemented, or other contextual factors.

3. *Ex-post* evaluation

Evaluation is a process that systematically and objectively assesses all the elements of a policy or programme intervention (e.g., design, implementation and results achieved) to determine its overall impact. The objective is to provide evidence-based information with a view to delivering ever-increasing results. There are two main types of *ex-post* evaluation: performance evaluations, which has already been discussed in the previous section, and impact evaluations.

The aim of the **impact (or effects) evaluation** of a public policy intervention is to check whether – and to what extent – the policy instruments put into place have served to bring about the intended changes. The main difficulty posed by this type of evaluation is to distinguish how much of the observed change is directly attributable to the public intervention and how much to other factors that can, in any case, influence change¹. Various approaches can be adopted to measure the effects of a policy intervention. The choice depends on the particular aspects of the intervention being evaluated and the specific evaluation questions that need to be answered. The two best established approaches in impact evaluation are²:

1. *Counterfactual impact evaluation*, which answers the questions “Did the planned results occur? Were the effects obtained caused by the intervention?”. This analysis focuses on the difference between the outcomes observed under the intervention and the hypothetical (“counterfactual”) situation that would have occurred in the absence of the intervention. This approach attempts to separate the changes caused by the specific intervention from all the changes that occurred – over a certain period of time – in the group or area affected by the intervention. The counterfactual method is suitable for evaluating limited interventions with clear objectives and precise implementation mechanisms³.
2. *Theory-based impact evaluation*, which analyses various complex situations caused by the different contexts and mechanisms triggered by an intervention. This approach answers the questions “What happened? What worked best, where, for whom and why?” by conceptualising the intervention in terms of a theory that explains how it works and which mechanisms led to the observed outcome.

Since the 2014-2020 programming period, the European Social Fund (ESF) provides for the use of impact evaluations for interventions co-financed through European resources. The ESF guidance document suggests that the two approaches outlined above should be used in a complementary way, in order to provide useful information for programming interventions in the following period⁴.

¹ P.H. Rossi, M.W. Lipsey, H.E. Freeman: *Evaluation. A systematic approach* (7th Edition, Sage publication, 2020).

² In addition to the “counterfactual” and “theory-based” approaches, social science also uses the “goal-free” approach (based on quality criteria and the needs of the people targeted by the intervention) and the “social process/constructivist” approach (based on explaining possible conflicts between groups of participants in order to arrive at more accurate and shared knowledge). These approaches and related methods can be found in G. Marchesi, L. Tagle, B. Befani: *Approcci alla valutazione degli effetti delle politiche di sviluppo regionale* (Ministry for Economic Development, Public Investment Evaluation Unit, 2011).

³ P.H. Rossi, M.W. Lipsey, H.E. Freeman: *Evaluation. A systematic approach*, *ibid.*.

⁴ European Commission: *Monitoring and evaluation of European cohesion policy. European Social Fund guidance document* (June 2015)

4. Counterfactual evaluation of the effects of public policies

The purpose of counterfactual impact evaluation is to assess a public policy's ability to change the behaviour or conditions of a certain target group in the desired direction. The effect is defined as the difference between what happened *after* the implementation of a policy (factual situation) and what would have happened if that policy had not been implemented at all (counterfactual situation). For example, the counterfactual evaluation of an intervention aimed at reducing worker exploitation in agriculture will measure the number of workers who are no longer exploited after the intervention compared to the number of the same workers who would no longer be exploited if the intervention had not been implemented. The difference between the two values is the net effect of the intervention.

The main difficulty in this approach is estimating the second term of the equation (how many of the same workers would have emerged from being exploited without the intervention), which cannot be directly observed. Hence the need to produce a reliable estimate of the effect by reconstructing the counterfactual situation with observable data to approximate what would have happened to the beneficiaries in the absence of the intervention⁵. These estimates can be made by using “control groups”, which are, for all intents and purposes, identical to the group under evaluation. If the two groups differ in some substantial way there is a “selection bias”. This means that the observed results at the end of the policy intervention can be due to these differences rather than to the implemented intervention.

This evaluation approach is particularly suited to well-defined interventions, which are uniformly applied in all cases and to all recipients/users, according to random chains that establish a linear link between the intervention and its effect. In addition, it should be ensured that:

- the *outcome variables* can be identified, to verify the presence of an effect using the available analytical tools and resources;
- *the intervention is clearly defined and identifiable*. Where policy components cannot be clearly identified as individual interventions, the overall scope of the evaluation is impaired. The treatment must be identifiable in its characteristics and in the consequences, it is expected to produce, and it must also be uniformly applicable to all beneficiaries.

In its quantitative version, the counterfactual approach uses two different methods called *experimental* and *non-experimental designs*.

4.1. Counterfactual evaluation: experimental designs

The basic idea of the experimental method stems from the clinical trials used to test the effectiveness of drugs. A number of patients suffering from a disease are divided into two groups by drawing lots (or random assignment). One group (experimental or ‘treatment’ group) is given the drug, the other (‘control group’) an inert substance. The progression of the disease observed among the members of the control group will thus be very similar to the progression

⁵ G. Marchesi, L. Tagle, B. Befani: *Approcci alla valutazione degli effetti delle politiche di sviluppo regionale*, ibid.

that the disease would have had among the members of the experimental group in the absence of the drug.

Although the experimental design (randomised experimental design or randomised controlled trial) offers the highest internal validity for assessing an intervention, the method is often unusable. For example, a randomised experimental design cannot be used when: (i) the intervention is already under way and the participants have already been selected by other methods; (ii) there are sufficient resources to have all eligible persons participate in the intervention and it would be unethical to deny a service or benefit for evaluation purposes only; and (iii) the intervention concerns a limited number of persons with unique characteristics⁶. When the experimental design is unviable, counterfactual impact evaluation relies on quasi-experimental designs.

The main problem with the experimental method is maintaining the integrity of the experiment, i.e., the clear separation of the control group from the experimental group for the entire duration of the experiment. On the one hand, there is the problem of non-participation and incomplete participation ('no-shows' and 'drop-outs') by those assigned to the treatment group; on the other hand, there is the problem of 'cross-overs', i.e. individuals assigned to the control group who still manage to benefit from the intervention. These are the main reasons for bias of estimates, which make a simple comparison between participants and non-participants impracticable.

The strength of this evaluation method lies in the immediacy of the notion of effect as the difference between observed and counterfactual. However, the difficulties of application are such as to severely restrict its practical relevance, even in those countries where the method is widely used (United Kingdom, Canada and the United States of America). Randomisation reduces the scope of applicability of this approach to the relatively rare situations in which the method is not hampered by ethical or political reasons, or by difficulties in obtaining the cooperation of service providers. The applicability of the experimental method is also limited to interventions with partial coverage. All interventions of a universal nature (such as primary education, environmental protection or the promotion of decent work), do not by definition lend themselves to the identification of a control group and thus to randomisation.

While randomisation offers the best method to ensure the internal validity of the estimates, the greatest limitation is found in their external validity. In particular, effects estimated on a small scale (typical of the experimental approach) do not take into account the constraints that arise when the intervention is applied on a large scale⁷.

⁶ ILO: *Monitoring the performance and evaluating the impact of the Youth Guarantee: A learning package*, ibid.

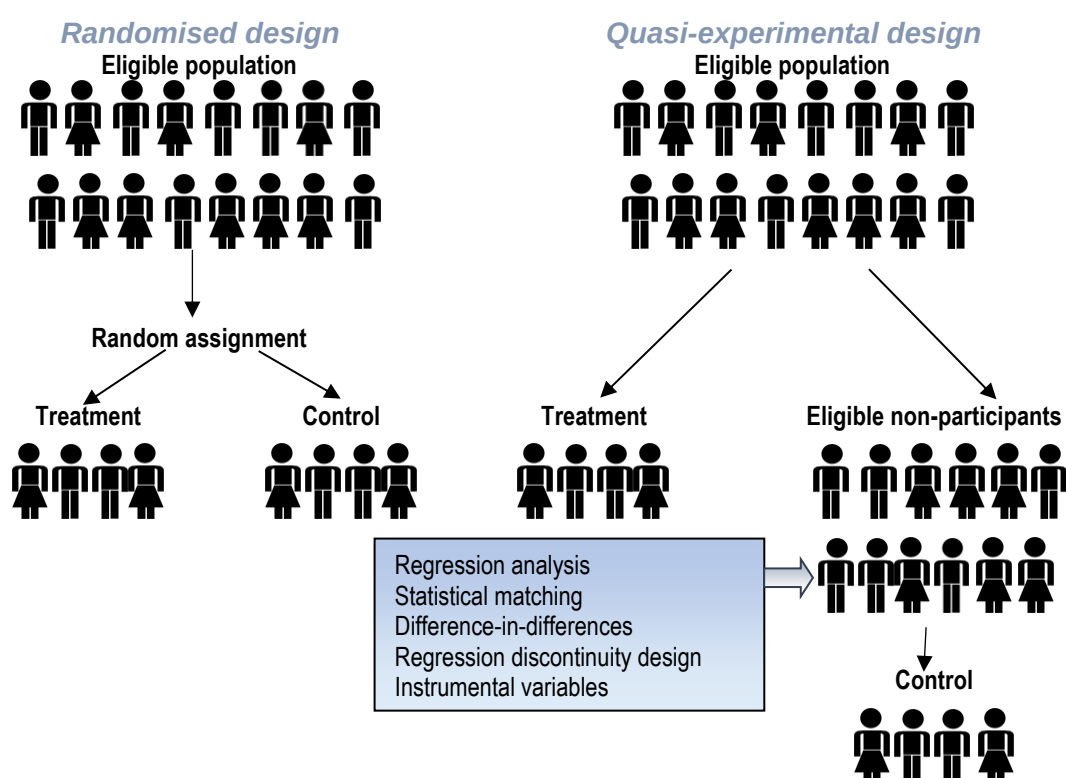
⁷ G. Marchesi, L. Tagle, B. Befani: *Approcci alla valutazione degli effetti delle politiche di sviluppo regionale*, ibid.

4.2. Counterfactual evaluation: non-experimental (or quasi-experimental) designs

Non-experimental (or quasi-experimental) methods are the most widespread, especially in Europe. They are used in all situations where the random assignment of the recipients of an intervention proves impossible. In these cases, the counterfactual can only be reconstructed by making assumptions which then have to be verified, depending on the quantity and quality of the available data.

At the core of quasi-experimental approaches are the different statistical methods used to build a valid control group (difference-in-differences, regression analysis, statistical matching, regression discontinuity design and the use of instrumental variables)⁸. The following graph illustrates the differences between randomised experimental design and quasi-experimental design.

Diagram 1. Randomised experimental and non-experimental approaches



Fonte: ILO: *Monitoring and evaluation of youth employment programmes* (ILO, Geneva, 2003)

The **difference-in-differences** (DD) method is commonly used in impact evaluations as an alternative to randomisation. The DD method is based on the concept that the difference between the post-intervention outcomes of the participants and the control group includes both the “real” treatment effect and the selection bias due to the differences between the two groups.

⁸ For a more detailed analysis of the various quasi-experimental techniques see G. Marchesi, L. Tagle, B. Befani: *Approcci alla valutazione degli effetti delle politiche di sviluppo regionale*, ibid. For a practical example of what to do with the various techniques, see ILO: *Monitoring and evaluation of youth employment programmes: A learning package* (ILO, Geneva, 2013).

This bias can be partially corrected if data on the outcome variable can be obtained for both groups in at least two different periods. Subtracting the difference between the two groups measured in an earlier period can remove the part of bias related to the different characteristics between the two groups that do not change over time. If the difference between recipients and non-recipients does not change over time, the difference-in-differences method eliminates the selection bias and produces correct estimates of the effect of the intervention. For example, applying the DD method to an intervention that offers public incentives to businesses to increase job creation measures the average number of employees of the businesses to which the resources are allocated and the employees of the businesses that receive no resources, one year after the allocation. This difference does not reveal the effects of the incentives because it is likely that the group of incentive-recipient businesses would have behaved differently from the group of non-recipient businesses with or without the incentives (the different behaviour is the selection bias). If the differences between the two groups of businesses observed in two periods of time were to remain constant, they could be subtracted from the final outcome to determine the net effect value⁹. The applicability of this method requires that the outcome variable can be measured repeatedly in different periods, among both the recipients and non-recipients. Another necessary condition is that the policy under evaluation is identifiable with a binary treatment variable, i.e., that there are treated and non-treated individuals. If there are partial or non-uniform treatments the method is inapplicable.

Statistical matching creates an *ex-post* control group of non-treated individuals who are as similar as possible in terms of observable characteristics (sex, age group, level of education, labour market status) to the participants (a non-treated “match” is found for each participant in the intervention). Once the control group has been identified, the treatment effect is the difference between the average outcome variable of the participant group and of the matched non-treated group.

Statistical matching (and regression) methods are based on the assumption of the “selection of observable variables”, according to which the variables that affect the selection process are observable and can thus be used to remove the differences in the starting positions. Where the selection of observable variables is not possible, **instrumental variables** can be used. The requirement is that there is an external factor that can influence the selection process, affecting the decisions of the individuals involved, and that this external factor has no effect on the outcome variable. The strategy consists in “breaking down” the treatment variable into two parts: a part determined by the external factor beyond the individual's sphere of control and a part determined by the individual's decisions related to unobservable variables. Thus, only the part which does not depend on the unobservable variables is used as treatment (in econometrics this external factor is called the ‘instrumental variable’).

The main advantage of the method of instrumental variables is to exploit situations similar to randomisation and thus obtain effect estimates with a high degree of reliability, whereas the main difficulty is to find or build such variables, which, by definition, must affect the outcome variable and must be independent of the treatment¹⁰.

⁹ ILO: *Monitoring the performance and evaluating the impact of the Youth Guarantee: A learning package*, ibid.

¹⁰ G. Marchesi, L. Tagle, B. Befani: *Approcci alla valutazione degli effetti delle politiche di sviluppo regionale*, ibid.

5. Theory-based evaluation

Theory-based impact evaluation is used for the causal attribution of an effect by conceptualising the intervention on the basis of a theory that identifies the mechanisms leading to a certain outcome.

According to this approach, it is assumed that interventions always have an effect (whether expected or unexpected, positive or negative) and always trigger mechanisms in the actor categories involved. If the actors are the recipients of the interventions, we speak of “programme theory”. If, instead, the actors are part of the chain ranging from the decision to the implementation of the intervention (decision makers, officials, managers, operators), then we speak of “implementation theory”. The two theories are linked because the resources produced during implementation (economic, social, cognitive, cultural, informative, relational, etc.) complement the pre-existing resources of the target group and influence its behaviour.

Theory-based evaluation is suitable in relation to investigations into the link between different forms of empirical evidence and to building the “causal chain”, which operates in depth and explains the theory behind a correlation or variation in an indicator. Typical questions in theory-based evaluation are “What exactly ensures that a given intervention will work? Which actors’ actions are decisive? What are the consequences of certain interactions and/or organisational arrangements? What do the actors involved in the implementation of a given intervention need in order to adopt a specific behaviour? How is the behaviour of the target group related to the actions taken by the planners?”. In the field of social sciences, theory-based evaluations mainly adopt the realistic approach.

5.1. The realistic approach to evaluating effects

The realistic approach proposes to explain an outcome through a mechanism and context. The ‘outcome = mechanism + context’ configuration is the key concept of realistic evaluation. The term “theory” here has a double meaning which includes both scientific theories, based on previous research, the results of which can be found in the literature, and the ideas elaborated by the actors involved (the implementers and beneficiaries of the intervention) about why an intervention works. All theories, irrespective of their source, should then be subjected to empirical research, i.e., they should be tested with the (qualitative and/or quantitative) methods most appropriate to the content of the theory¹¹. Evaluation based on the realistic approach comprises four steps¹²:

1. Through dialogue with programme implementers, the evaluator tries to understand the nature of the intervention: What is the purpose of the programme? What is the nature of the target population of the programme? In which contexts and settings does it operate? What are the prevailing theories about why the programme works for some people in some circumstances?

¹¹ R. Pawson, N. Tilley, ‘Introduction to realistic scientific evaluation’, in N. Stame N. (ed.), *Classici della valutazione* (Franco Angeli, Milan, 2007).

¹² Cf. A. Blamey, M. Mackenzie “Theories of change and realistic evaluation: Peas in a pod or apples and oranges.” in *Evaluation*, N. 13(4), 2007, p. 439-455.

2. The evaluator draws a series of potential mini theories relating the various contexts of an intervention to the multiple mechanisms through which it might operate.
3. In the third phase, the evaluator conducts an outcome enquiry into these mini-theories. This involves the development of a quantitative and qualitative model of the intervention.
4. Therefore, the analysis of how context-mechanism-outcome configurations are implemented within a programme allows the evaluator to refine the theories about what works, for whom and in what circumstances.

Box two provides a practical example of a context-mechanism-outcome configuration used for the realistic evaluation of income support policies.

Box 2: Context-mechanism-outcome relationship in social benefits

Social benefits provided at local level to people at risk of poverty are successful if they can offer them the means to become independent from public financial support (*outcome*). In some cases, social services achieve this goal because they offer complementary measures, in addition to the benefits, which strengthen the proactive attitude of the beneficiary and increase his/her efforts to look for work (*mechanism*). In other cases, the positive outcome may be due to the fact that the service offered expands the resources available to the recipient, who therefore can increase control over his/her situation (*mechanism*), e.g., by offering childcare services to single mothers, in addition to a welfare benefit (*context*). Single mothers, by gaining a better control over their situation, can thus pursue medium- to long-term strategies (*mechanism*), such as training, which in the long run will enable them to climb out of the welfare system, which would prove a difficult task if childcare fell entirely on their shoulders without any additional benefits (*context*).

Instead, in cases where the support measure (childcare service) is not provided on a regular basis (*context*), the women benefiting from this intervention do not see it as a resource on which they can permanently rely and are therefore discouraged from adopting medium/long term strategies (*mechanism*). Other cases of failure occur when the labour market can only offer precarious or irregular jobs (*context*), thus inducing recipients to prefer the low but certain and regular income from benefits, which can even become a supplement to their main income (from informal work) (*mechanism*).

Context	Mechanism	Outcome
Social services offer supplementary measures on top of benefits	These measures encourage the development of a proactive attitude in people and enhance their job-seeking efforts	Success
Single mothers are offered childcare services alongside the benefits	By increasing control over their own circumstances single mothers can elaborate long-term strategies for climbing out of welfare	Success
A welfare measure is irregular and therefore perceived as unreliable by the beneficiaries	Beneficiaries are discouraged from elaborating medium-term strategies to improve their circumstances, preferring short-term strategies instead	Failure
Job offers provide an uncertain and irregular income	Beneficiaries prefer the low but certain and regular income from welfare benefits	Failure

Source: F. Biolcati Rinaldi, *Povert , teoria e tempo. La valutazione delle politiche di sostegno al reddito* (Franco Angeli, Milano, 2006)

The realistic approach can be used when counterfactual impact evaluation is unviable (due to different implementation in different geographical areas, lack of implementation uniformity) or when the factors that could influence the outcome are unclear.

In the context of social change typical of public policies, actions are differentiated, strongly linked to the contexts in which they take place and complex (numerous actors, non-

linear causal relationships, unclear interactions between the elements of the intervention). It is therefore difficult to isolate the factors influencing the outcome and to use them as uniform statistical units – not only because of the different contexts and mechanisms, but also because of the diversity of outcomes, which vary from community to community¹³. The value system and socio-economic level have an important influence in defining both the specific characteristics of the intervention (and thus the specific and operational objectives) and the way the beneficiaries react. In these cases, realistic evaluation can be useful because it asks not only if the intervention works and to which extent, but also what works about the intervention, and in particular what works best, where, for whom and in which circumstances. A possible framework for conducting a realistic evaluation is outlined in Box 3¹⁴.

Box 3: Outline of a realistic evaluation of the effects of an intervention

Defining the purpose of valuation	Identifying the evaluation question	- What is the nature and content of the intervention? - What are the circumstances and context? - What are the policy objectives? - What is the nature and form of the results? - Conducting exploratory research to inform discussion with decision-makers
	Defining the purpose of evaluation	- Integrity of theory - does the intervention work as intended? - Choice of theory - which intervention theories seem to fit best? - Comparison - how does the intervention work in different environments and for different groups? - Testing - how does the policy intent translate into intervention practice?
	Researching and defining the intervention theories	- Searching for explanatory theories in the literature - Grouping, categorising and synthesising theories emerging from the literature - Define an evaluation framework to be populated with empirical evidence

¹³ G. Marchesi, L. Tagle, B. Befani: *Approcci alla valutazione degli effetti delle politiche di sviluppo regionale*, ibid.

¹⁴ R. Pawson et al.: *Realist synthesis: an introduction* (ESRC Research Methods Programme, University of Manchester, Paper 2/2004).

Box 3: Outline of a realistic evaluation of the effects of an intervention (cont.)

Searching for and evaluating evidence	Searching for evidence	▪ Defining the sampling strategy ▪ Defining the research sources, terms and methods to be used
	Evaluating evidence	▪ Relevance of tests - does the research address the theory under investigation? ▪ Rigour of tests - does the research support the conclusions drawn by the evaluators?
Extracting and synthesising results	Extracting results	▪ Developing a data collection system ▪ Gathering data to populate the scoreboard
	Synthesising results	▪ Comparing the results with those in the literature (both positive and negative) ▪ Refining intervention theories in the light of the empirical evidence gathered
Conclusions and recommendations		▪ Involving decision-makers in the analysis of results ▪ Elaborating and verifying conclusions and recommendations with key actors ▪ Sharing results, conclusions and recommendations

Source: Adaptation based on R. Pawson et al.: *Realist synthesis: An introduction* (ESRC Research Methods Programme, University of Manchester, Paper 2/2004).

Biography

- Biolcati Rinaldi F., *Povert , teoria e tempo. La valutazione delle politiche di sostegno al reddito* (Franco Angeli, Milano, 2006)
- Blamey A., Mackenzie M. “Theories of change and realistic evaluation: Peas in a pod or apples and oranges.” in *Evaluation*, N. 13(4), 2007, p. 439–455
- European Commission: *Monitoring and evaluation of European cohesion policy. European Social Fund guidance document* (June 2015)
- Fujiwara D. 2010. *The Department for Work and Pensions Social Cost-Benefit Analysis Framework* (Department for Work and Pensions working paper No. 86)
- Gertel, P.J. Martinez, S., et al. 2011. *Impact evaluation in practice* (Washington, DC, World Bank)
- Marchesi G., Tagle L., Befani B.: *Approcci alla valutazione degli effetti delle politiche di sviluppo regionale* (Ministero dello sviluppo economico, Unit  di valutazione degli investimenti pubblici, 2011)
- Mason P., Barnes M.: “Constructing Theories of Change”, in *Evaluation* Vol 13(2), 2007 p. 151 – 170
- New Economics Foundation (NEF). 2013. *Social CBA and SROI* (Economics in policy-making, Briefing note 4)
- International Labour Organization (ILO): *Monitoring and evaluation of youth employment programmes: A learning guide* (ILO, Geneva, 2013)
- *Monitoring the performance and evaluating the impact of the Youth Guarantee: A learning package* (ILO, Geneva, 2017)
- *Social cost-benefit analysis of the Youth Guarantee: A learning package* (ILO, Geneva, 2017)
- Pawson R. et al.: *Realist synthesis: An introduction* (ESRC Research Methods Programme, University of Manchester, Paper 2/2004)
- Pawson R., Tilley N., “Introduzione alla valutazione scientifica realista”, in N. Stame N. (ed.), *Classici della valutazione* (Franco Angeli, Milano, 2007).
- Rogers, P. (2007). Theory-based evaluations: Reflections ten years on. *New Directions for Evaluation*, 114, 63–67
- Rossi P.H., Lipsey M.W., Freeman H.E.: *Evaluation. A systematic approach* (7th Edition, Sage publication, 2020)
- Stame, N.: “Theory-based evaluation and types of complexity” in *Evaluation*, 10(1), 2004, p. 58–76