

Assess whether a programme or policy makes a difference

Typical errors

Too often evaluations are commissioned at a single point in time, with little or no baseline data. They therefore only give a static picture of a programme/policy at a given point in time. While this is better than no data, it does not enable evaluators and policy makers to make judgements about:

- what change can be attributed to the programme, and subsequently
- whether the programme/policy is successful or not.

A good quality evaluation should capture change and evolution not just as a one-off picture against key indicators. Typically, a time span of several years is needed to capture the final impacts. However, a mid-term evaluation may be needed earlier to verify that the programme is 'on track' in delivering the outputs and intermediate results expected. If the time span is too long and there is no ongoing monitoring, it will not be possible to capture many of the immediate outputs and short-term results.

A second recurrent issue in programme evaluation is that if change is identified, it is assumed to be due to the programme/policy. However, many other factors could have contributed to the change. This issue is referred to as 'attribution' - i.e. can the identified change be attributed to the implemented policy, or would it have happened anyway?

The attribution of change to a programme should therefore be assessed rather than assumed.

A third problem is overlooking the implementation process. The implementation process very often explains why performance is not as good as expected, or why there are important variations between regions. If the evaluation focuses on the results and impacts alone, there is a risk that it will provide little insight into how policy implementation should change in order for the results to improve.

Measuring change

There are two main approaches to capturing change:

- **Macro and meso-level indicators:** comparison with baseline data including the option of a longitudinal evaluation design. For example, the NEET rate in a region before the introduction of an intervention is compared with the NEET rate one, two, three, etc. years after the introduction.
- **Micro-level indicators:** pre - post measurement of beneficiaries. The situation of the same group of young people is assessed before and after they take part in a programme. The after measurement can be done at several points in time to assess if the effect remains: at exit, one year later, two years later, etc. Examples of indicators which could be measured through this approach are: average length of time for beneficiaries to find employment; average annual working time per employed project beneficiary.

When measuring change, you need to reflect on:

- Is there monitoring data you can use as baseline?
- Is it possible to follow the same group of young people over a period of time?

Attribution

The fact that change in key indicators is identified does not necessarily mean that the change is due to the programme/measure being evaluated. It is possible that the evolution would have happened even without the programme. Examples of possible factors influencing change which are not linked to a programme are:

- The context has changed: e.g. there is improved access to childcare services, and young mothers can (re)enter the labour force.
- Participants are taking part in more than one support programme: e.g. young people may be participating in an upskilling programme that improves their IT skills and at the same time they may take part in a career fair focused on jobs in the IT industry.
- The group of selected participants has some characteristics which explain the change in indicators: e.g. they are highly motivated at entry which explains positive outcomes which would not be found if the group was less motivated. For example, quite often pilot projects report more positive results than when the initiative is scaled up. Pilot initiatives typically attract the most motivated organisations (firms for example) and the high motivation of staff has positive effects on the results. Once all firms in a sector are asked to take part, this important factor is no longer present, and the results are less positive.

There is always a risk that the change is not caused by the programme being evaluated. If that is the case, an evaluation that does not verify attribution could conclude very positive findings and recommend higher investment in an ineffective programme. This would not be cost-effective.

In order to know whether a change in indicators is due to the programme evaluated, it is important to use the appropriate method.

When measuring change, you need to reflect on:

- What else besides your programme/policy could have caused the observed changes?

Measuring attribution

Attribution can be assessed both quantitatively and qualitatively.

Quantitative designs can be:

- experimental, or
- quasi-experimental

In experimental designs the group of potential beneficiaries is randomly allocated to the treatment group (those who take part in the programme) or a control group (those who do not take part in the programme). This guarantees that other than the programme, there are no or very few differences between the groups which could explain differences in outcomes. An example of an experimental design could be to randomly decide which NEETs will receive mentoring and which will not, and to measure the difference in results. This has a number of practical as well as ethical implications and would need a very carefully constructed research design.

Quasi-experimental designs also compare the outcomes of a treatment group and a control group, but without the randomised selection. For example, a treatment group could be defined as inactive young people who take part in a regional measure providing mentoring

and coaching to reintegrate them into the labour market. This support is additional to other active labour market policies. A comparable control group could be designed from another neighbouring region where inactive young people do not receive mentoring or coaching, and only take part in the standard active labour market policies.

The challenge in these studies is to define a control group which is truly comparable to the treatment group. For example, it would not be correct to compare participants who receive support because they are inactive, with all young NEETs in a different region – including those actively looking for a job. There are various techniques that can be used to ensure that the control group and treatment group are sufficiently comparable.

Both experimental and quasi-experimental designs can yield robust findings, but there are constraints on the use of both. They are therefore not always feasible or not necessarily the most efficient (in terms of resources needed).

Judgements on the attribution of the programme can also be made based on non-experimental approaches. Contribution analysis is one such technique. It is based on an explicit programme theory which spells out exactly how and why a programme is expected to lead to positive results. The programme theory also identifies the conditions under which the programme is expected to work. The evaluators assess the extent to which the programme theory holds, based on qualitative interviews with a range of persons – including beneficiaries but also those delivering the programme. These interviews either support the theory (and explain positive outcomes) or show gaps in the theory which explain why the outcomes are less positive than expected.

Non-experimental approaches are also useful to find out what went wrong when there are no changes in key indicators.

Note that there are different schools of thought on the merits of each of these approaches which are not discussed here.

Process evaluation

Some policies are fairly simple, and their implementation process does not have many variations. For example, if the policy provides a voucher (say €50) to all young people who take a skills evaluation test, there would be very little scope for varying the implementation.

However, most policies tackling the NEET problem do not fall into this category. They combine multiple activities and incentives, and the quality of the delivery very much depends on the people in charge and the institutional conditions in which they operate.

The same programme can be delivered with great success in one youth centre and be a failure in another, just because of the different process and institutional conditions.

This is why evaluations should not only assess the results and impacts, but also analyse the implementation process. Without looking at the process, there is a risk of missing important messages for future improvements.

Process evaluations assess whether the programme/policy activities have been implemented as intended. They also assess the barriers and success factors for implementation.

Delivery of a mentoring scheme, for example, can be subject to huge variations. Issues such as what is the profile of the mentors, how qualified are they, what is the quality of their working life, what mentoring methodology they follow and whether they actually follow that methodology can influence the results.

