Assessing whether the programme or policy makes a difference

Introduction

If we read in an evaluation report on the implementation of a support measure that the completion rate in a certain programme is 80%, what does this tell us?

This information alone does not tell us much. We would need to know what the completion rate was before the support measure was implemented.

Read below more about the typical errors made in evaluations.

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 really 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 a one off picture against key indicators. Typically a time span of several years would be needed to capture the final impacts. However a mid-term evaluation may be needed earlier to verify that the programme is ‘on track’ and 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 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 change measured be attributed to the policy implemented, or would it have happened anyway?

The attribution of a 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 schools/regions. If the evaluation focuses on the results and impacts alone, there is a risk that it will provide little insight into how the 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 rate of early leaving in a region before the introduction of an intervention is compared with the rate one, two, three, etc. years after introduction. Another example would be the share of early leavers from a given school before the school took part in a programme and after.
- **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 measure. 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: attitude to education and training, future aspirations, relationship with teachers.

When measuring change, you need to reflect on:

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

Attribution

The fact that change in key indicators is noted does not mean that the change is necessarily due to the programme/measure being evaluated. It is possible that the evolution would have happened anyway, 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 are less opportunities for unqualified young people on the informal labour market and the 'pull factor' is no longer present. They are therefore not tempted to quit school for a job.
- Participants are taking part in more than one support programme: e.g. early leavers receive remedial training and at the same time they are in a programme to tackle youth unemployment and which aims to get them primarily into a job (independent of the qualification).
- The group of participants selected 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 replicated on a large scale. Pilot initiatives typically attract the most motivated organisations (schools for example) and the motivation of staff has positive effects on results. Once all schools 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 due to the programme. If that is the case, an evaluation that does not verify attribution could conclude very positive findings and recommend higher investment in the programme. This would not be a cost-effective decision.

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 students 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 unemployed early leavers who take part in a regional measure providing mentoring and coaching to reintegrate them into education. This support is additional to other active labour market policies. A comparable control group could be designed from another neighbouring region where unemployed early leavers 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 at risk of early leaving, with all students in a different school – including those not at risk. 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 very 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).

Experimental and quasi-experimental approaches alone typically focus on the results and impacts. They provide very little insight into the implementation process and its strengths and weaknesses.

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.

Examples on measuring attribution:

The French intervention programme called Parent’s toolbox was evaluated using an experimental design. In 40 participating schools, the evaluators selected at random 100 classes which implemented the methodology of parents’ toolbox (the students in these classes constitute the treatment group). The other classes did not implement this methodology and changed nothing compared to previous year. They constitute the control group.

2010 evaluation report (in French)

The effectiveness study on the Dutch measure Medical Advice for Sick-reported Students (MASS) used a quasi-experimental design. 7 out of 21 schools for pre-vocational secondary education had been applying the intervention programme, and were all asked to participate

in the study (intervention schools). Within the group of remaining schools, seven were asked to participate in the study as control schools. Control schools were selected by their characteristics to match as much as possible those of intervention schools in terms of urbanisation, fields of education, and size of the school.

2016 study (in English)

Process evaluation

Some policies are fairly simple for which the implementation processes do not have many variations. For example, if the policy was to provide a voucher (say €500) to all students who successfully qualify without any additional support, there would be very little scope for varying the implementation.

Most policies to tackle early leaving do not fall into this category. They combine a number of activities 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 school and be a failure in another, just because of the process and the institutional conditions.

That 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 overseeing 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 really follow a methodology can influence the results.

Example:

The evaluation of the Austrian Youth Coaching Scheme analyses several aspects of its implementation. For instance, it gathers information on the working conditions of coaches, their qualifications, their linguistic knowledge, and their participation in team meetings and activities for continuing professional development.

2013 evaluation report (in German)

Source URL: https://www.cedefop.europa.eu/en/en/tools/vet-toolkit-tackling-early-leaving/evaluate/assessing-whether-programme-or-policy-makes-difference