

Evidence and Evaluation

Position paper of the DeGEval – Gesellschaft für Evaluation)¹

At its annual convention in 2012, the *DeGEval - Gesellschaft für Evaluation* discussed the topic “Evidence and Evaluation”. At present, this issue is quite significant in policy and practise settings. Decisions aimed at introducing and implementing political programmes should be based on knowledge. The connection between evidence and evaluation is self-evident since the latter is to generate essential information for the rational design and control of programmes and organisations. Those in charge of decision making and execution of evaluations expect findings based on scientific research in order to further support their own action. Whereas “evident” frequently refers to “self-evident” or “obvious” in everyday use, the term “evidence” in an evaluation context means “proof” or “argument”, so that knowledge based and well-founded decisions can be made.

Until only a few years ago, talks preceding evaluation projects took place on a relatively simple level, and they were often limited to questions like how to explain to the persons who are to be evaluated what an evaluation is or how to explain to contracting bodies the respective advantages of a specific evaluation. Nowadays, however, evaluators increasingly come across a well-developed understanding and a sound knowledge with regard to evaluations. Meanwhile, a certain amount of maturity as to the handling of evaluations seems to be widely spread.

In the past, evaluations covered a wide range of different issues: they examined the basic requirements for projects or programmes, they assessed their structure and meaningfulness, and there was a strong interest in accompanying measures by means of evaluation and in continuously developing them in cooperation

¹ Position papers contain short and generally understandable statements or opinions with regard to topical evaluation issues. They are prepared by the board of the “Gesellschaft für Evaluation”.

with the persons to be evaluated. This went hand in hand with the application of a large variety of methods.

Such a wide thematic and methodological range of evaluations has now been reduced in the face of an increasing (cost) pressure on initiators and executors of political programmes. This means that at present evaluations are more and more expected to supply evidences in the form of sound proof to demonstrate the efficiency of interventions. The occurrence of desired effects should be proven as clearly as possible.

Within the methodological debates in social sciences, the idea of evidences being able to support the efficiency of programmes solely by applying the methodical “gold standard” is dominating the respective discussions at present. Expressed in a simplified manner it means that effects can only be proven when a systematic comparison between at least two statistically identical groups, one with and the other one without “intervention”, results in a significant difference. The intervention must therefore be controlled, i.e. it must be carried out in an exactly prescribed manner and must therefore be repeatable.

However, such a randomised and controlled experiment is difficult to realise in social contexts for a number of reasons. In social or political programmes, the variables influencing the success are scarcely isolable. Programmes can usually not be carried out in a strictly mechanical sense, and a great deal of programmes lack a logical causal model. For ethical and practical reasons, it is hardly possible to determine control groups, and quantitative studies require large samples.

Furthermore, the question might be asked what kind of content related statements can be gathered by means of quantitative effect sizes of programmes. Merely looking at effects might easily shift the view to non-intended or even negative effects of programmes. This applies in particular to underfinanced evaluations.

The expectation of effect evidences is quite understandable. Indeed, effect evaluations used to be an essential part of the evaluators’ work. However, the dominance of requests for effect evidences tends to overlook the fact that the quality of a programme can neither be sufficiently assessed nor further developed by effect measuring. Such a shift or narrowing of focus increasingly alters the function of an evaluation: it is not so much aimed at optimisation but at legitimising a programme or a measure.

Let us look at the problem from another perspective. If an evaluation has found proven evidences, it is certainly disappointing when the decision maker or the body in charge of practical implementation will then not act according to the evidences provided. Those who evaluate must learn to understand that decisions can well be supported by the respective evidences, but that they are still influenced by other variables as well. Evidences do not determine certain decisions. The way one acts always depends on contexts, norms and standards that are beyond the scope of proof. Decisions might be based on information gathered on the basis of evidences, but they might still find orientation in other standards and benchmarks than measured evidences. Action in political, social and educational contexts cannot be reduced to “something that can be measured” in the sense of a normatively neutral currency.

During the exchange between evaluating body and awarding body, the plausible request for generating evidences should always bear in mind the difficulties involved with such a requirement. Evidences should also serve the purpose of improving programmes, increasing their advantages, and strengthening the people – and should not just find proof for measurable effects. This approach requires feedback loops and their analyses in the sense of structured learning processes. It also requires a multi-methodological procedure, which is not possible without adequate time and effort as well as the respective expenditure that goes hand in hand with it.

Ideas of straightforward and unproblematic effect evaluations are unrealistic and reduce the potential of evaluations. Professor Geert Biesta (at present at Luxembourg University), the main speaker at the DeGEval conference in 2012, emphasised the fact that the overpowering idea of evidences generates the tendency that decision makers only deem important what is actually measurable with regard to its effects. Instead of that, it should be quite the other way round. The main question should be what is important to us as far as political, societal, educational and ecological issues are concerned. The question if the important issue is in fact measurable should then actually be secondary.