
6. Social and political context of evaluation

Reinhard Stockmann

Context is everywhere. Wherever we are, whatever we do, *context* surrounds us and influences our life and the way we act. For that reason, contextual conditions and factors simply cannot be ignored – i.e., not in evaluations either. But what do we actually mean by context, and how can it be taken into account in evaluations? We aim to study this question here first before explaining that evaluations not only constitute a scientific-methodological activity, but are part of a political process (Chapter 6.1).

Stakeholders, in the broadest sense, are part of the context. First, we explain what is to be understood by this somewhat bulky term and which groups of stakeholders exist. It is a largely incontrovertible fact that the participation of the stakeholders is important to the successful implementation of an evaluation and the use of its results. It is just that there are some very different attitudes as regards the intensity and degree of participation. The *participative CEval model* is presented here as an example of a pragmatic approach (Chapter 6.2).

The requirements relating to evaluators working *internally* or *externally* differ in a number of points, which will be dealt with in detail in Chapter 6.3. The relationship between evaluators and stakeholders is not always cloudless. Since evaluations are integrated in a *political process*, they are not only assessed applying scientific and subject-specific criteria, but also according to their political usefulness and usability. But the evaluator is confronted with a series of challenges which can put him under a great deal of pressure. To prepare evaluators for that, a further section deals with potential attempts to influence evaluators and the types of criticism this may involve. So that the interaction processes between evaluators, clients and other stakeholders can evolve as fairly and respectfully as possible, evaluation societies and international organisations have adopted standards and ethical guidelines which can be used as aids to orientation (Chapter 6.4).

Evaluations are associated with a wide range of challenges and great expectations. That means that evaluators need comprehensive, wide-ranging *competences*. Chapter 6.5 explains what these are. It is true on the one hand that everyone agrees that an evaluation ought to be useful, but the question of exactly what is to be understood by that is a controversial one. The state of knowledge on the empirical *use* of evaluations, what kinds of use exist and what their dimensions are, what factors have an influence on the development of that use, and what usefulness standards there are, are topics dealt with in Chapter 6.6.

6.1 DEFINITIONS AND CONTEXT CONCEPTS

6.1.1 Context: An Amorphous Term

According to the *Cambridge Dictionary*, context is ‘the situation within which something exists or happens, and that can help to explain it’ (<https://dictionary.cambridge.org/dictionary/english/context>). Or elsewhere: ‘The surroundings, circumstances, environment, background

or settings that determine, specify, or clarify the meaning of an event or other source' (<https://www.yourdictionary.com/context>). Although these definitions are very general and abstract, so that context appears to mean practically anything, they do give us a decisive clue. Among the innumerable contextual conditions that exist, it is the ones that are relevant which can explain something or give an event a specific meaning.

With reference to evaluation, *context* is the environment in which the evaluation object – an intervention, programme, policy etc. – is located. So context comprises the social and political environment, the geographical location, the various different organisations, groups and individuals involved, and much more. What context actually means in an evaluation always depends on the specific situation. That is why any attempt to make a comprehensive definition of context is bound to fail on account of its amorphous nature. What context actually means in an evaluation, what is relevant and therefore needs to be taken into account, depends on the evaluation object. On closer examination, the context can turn out to be a decidedly complex structure; many evaluands, such as programmes, for example, are implemented in multiple contexts, which can cover various different social systems, sectors or political units, networks, organisational levels at which interventions take place, stakeholder groups etc.

Greene (2005, p. 82f.) attempts to categorise the complexity of this term with *five dimensions*. She distinguishes between:

- (1) demographic characteristics of the setting and people in it;
- (2) material and economic features (e.g. buildings, resources);
- (3) institutional and organisational climate in a setting (character of the implementing organisation and its culture);
- (4) interpersonal dimensions (interactions and norms that frame it);
- (5) political dynamics (in terms of interests, power, influence and privilege).

Even if this list does help to structure the evaluation context, it cannot, by its very nature, be complete. Other authors, for example, add other dimensions. The most influential one discussed in the evaluation literature is '*culture*', often linked to the requirement that evaluation must be 'responsive to different cultures' (cf. Conner 2005; Hood, Hopson and Frierson 2005; Chouinard and Cousins 2009; Askew, Beverly and Jay 2012; Hood, Hopson and Frierson 2015; Brown and Di Lallo 2020). Without doubt, culture is an important contextual dimension, and it is not only significant in cooperation with the global South, but in Germany too, where there are also many different cultural milieus. However, using the terms context and culture synonymously leads to an unacceptable narrowing of the term context which, as shown, can cover a great number of dimensions (cf. Fitzpatrick 2012, p. 9).

Even if attention is drawn again and again to the significance of context for evaluations, there have not yet been many attempts to identify important *context-related factors* systematically and include them in programme planning and evaluation processes. Debra Rog developed a 'context framework' for that, which comprises five areas: 'the nature of the particular phenomenon or problem, the nature of the intervention, the broader environment in which the intervention is set, the parameters of the evaluation itself, and the broader decision-making area' (Conner, Fitzpatrick and Rog 2012, p. 90; Rog 2012, p. 25ff.).

In each of these areas, seven dimensions should be investigated: physical, organisational, social, cultural, traditional, political and historical. Additionally, sub-dimensions such as gender, race or language, power, class, sociopolitical status and others are recommended.

Any evaluation practitioner can easily imagine what such an analysis would require in terms of human and time resources. Apart from that, these are context-sensitive evaluations, in which the 'background' becomes the 'foreground', not connected with further evaluation approaches, although in some of those evaluations the context will have been incorporated quite early on.

6.1.2 Context in Evaluation Approaches and Planning Procedures

Stufflebeam (1968, 1971) was one of the first to highlight context consciously, in his Context-Input-Process-Product (CIPP) model. Having said that, the context was mainly limited to a 'needs assessment'. Stake's responsive model (1980) also drew attention to the fact that the programme should be analysed in its context in order to recognise local needs, give the local stakeholders a voice and tap local know-how. Weiss (1972, p. 108) was one of the early authors to draw attention to the different 'worlds' in which evaluators, programme planners, users and other stakeholders live, and to the fact that they have their own values and kinds of behaviour etc., act in different kinds of network, pursue different interests etc. And that is why she calls for contextual elements to be taken into account in an evaluation.

A look at the evaluation approaches presented in Chapter 2 shows that each of them conceptualises context in a very different way. The management-oriented approaches (greatly simplified) focus mainly on the information requirements of the programme managers and the context of the organisations which are intended to use the evaluation findings, like Patton (1978, 2008), for example, with his 'utilisation-focused model'.

With their 'realistic evaluation' model, Pawson and Tilley (1997) propose a completely different procedure: the aim of the analysis of the context is to find out under what contextual conditions programmes succeed or fail. This aspect is particularly important for the generalisability of evaluation findings.

In the theory-oriented approaches too, contextual conditions are understood as causing variables, defining context 'as an inevitable and thus rich source of explanatory influences on desired program or performance outcomes' (Greene 2005, p. 83). That is to say that observed changes are explicitly put in relation to context factors to identify 'causal mechanisms'.

In order to be able to isolate individual causes, experiments or quasi-experiments are often used, in which the aim is to isolate the variables that are theoretically responsible for changes from other context variables, so as to be able to provide empirical evidence of the causality between the intervention and the impact. If it is not possible to randomise, as many of the context factors that could have an influence on the outcome of the intervention as possible are kept stable by conscious selection, ensuring that the intervention group and the comparison group (to which the measure does not pertain) resemble each other in as many parameters as possible, so that the intervention caused – for example, by a programme – makes the difference. Thus, if there is a difference in the result between the two groups after an intervention, it can be attributed to the intervention, given that all the other context factors were kept stable. In experiments with randomisation there is no doubt about the causal explanation, because the random selection ensures that the distinguishing characteristics are evenly distributed. In a quasi-experiment in which there is no randomisation, other distinguishing characteristics between the two groups, apart from the intervention, could also be responsible for differences observed in the result. This can happen when features of the context are not considered impor-

tant and thus not taken into account in the formation of the groups, although in fact they are important (cf. in detail Chapter 7).

The viewpoint of the participatively and assessment-oriented evaluation approaches, which often use qualitative methods, is completely different. In these, context is ‘not considered a separable influence on program experiences and outcomes but rather is viewed as an important constituent of them’ (Greene 2005, p. 83f.).

In other words, information can only be understood and interpreted at all in its context: ‘de-contextualised information loses its meaning’ (ibid.) In such evaluation approaches, therefore, particular importance is placed on a detailed description of the multiple dimensions of the context. In the assessment-oriented approaches, the fundamental values and norms of the various actors are to the fore.

In certain participatory or even empowerment approaches, the context factors power, influence, social inequality, justice etc. play an important role, as the central goal consists of changing such contextual conditions, as a result of which evaluation explicitly becomes a political process – a point that will be taken up again later on.

For the record, almost all the evaluation approaches – except the target-oriented ones, which are hardly used any more in their unadulterated form anyway – agree ‘that context matters’ (Greene 2005, p. 84), but they differ greatly about how context is to be treated and addressed. Moreover, hardly any of the prevalent evaluation approaches offer a theoretical rationale or even an analytical guide to how context in its various different dimensions is to be included systematically in a ‘context framework’ and analysed (as, for example, proposed by Conner, Fitzpatrick and Rog 2012; Rog 2012).

A look at the established *programme planning procedures* also shows how variously the construct of context is handled. A widely used planning instrument is the *logical framework approach*, referred to here in its abbreviated form as the Logframe. In the Logframe, individual activities are assumed, which generate certain outputs, which are in turn intended to evoke desired impacts (outcomes), which, at the end of the day, are intended to lead to the achievement of the aim of the programme. The Logframe contains a detailed description of how this ‘impact chain’ comes about, but it does not designate the causal relationships clearly. The aim is to measure the progress or success of the programme using indicators which can, moreover, be given target dimensions. The context does not have much of a role to play in this contemplation; instead, it is relegated to the assumptions or risks via a description of the conditions under which the expected outputs and outcomes will or will not come about.

Figure 6.1 shows such a Logframe example which makes it clear that, as ‘assumptions’ and ‘risks’, the contextual conditions get marginalised in the truest sense of the word.

The programme theory approaches, which focus on the *theory of change*, avail themselves of a completely different procedure.¹ The theory is not merely intended to describe why a programme functions under what (contextual) conditions, but also to explain the reasons. A precise analysis of the context of an intervention or programme is necessary for this. The core problem is defined as the deficient state of affairs in the context which is to change for the better. Inhibiting and promoting structures, processes and human behaviour – i.e., context

¹ Funnell and Rogers (2011); Rogers (2014a, 2014b); World Food Program (2017); Jones and Rosenberg (2018); Theory of Change – Guidance on Developing, Representing and Using Better Evaluation (https://www.betterevaluation.org/en/toc_goodpractice).

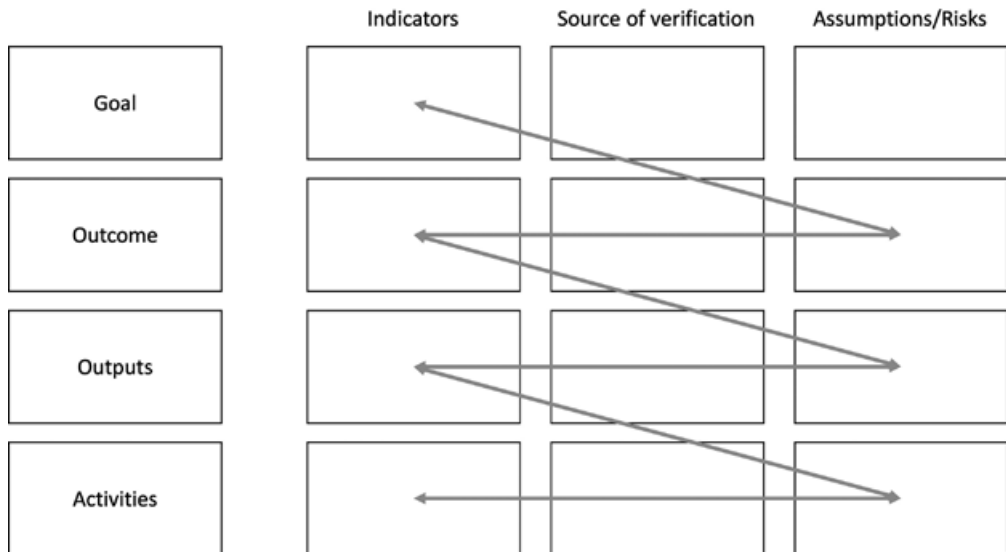


Figure 6.1 Logical framework structure

factors – are identified, to determine exactly where a programme should intervene so that the intended change comes about.

The main initial questions are: (1) What are the reasons for the existence of a deficient situation (for example, a situation in which the child mortality in a given ethnic region is much higher than the national average)?; (2) What needs to change in the contextual conditions so that the deficient situation changes for the better (for example, the construction of health centres, training of healthcare staff, construction of roads and transport infrastructure so that parents can access the health centres etc.)?; (3) How much of this can be influenced by the available resources and possible activities and how much cannot?; (4) Where exactly should the programme begin, and what should it focus on to make the deficient state of affairs disappear? (For example, to bring the child mortality in ethnic region X down to a rate that is just the same as the national average). Figure 6.2 features a simple form of a theory of change.

To sum up, it can be said that a theory of change is developed from the context. First, all the possible causes of a state of affairs recognised as being deficient are to be identified with the aid of studies, reports, other evaluations, stakeholder surveys etc., for example in a planning workshop. Then it is determined what would need to change to solve the problem, and finally there is a discussion of whether or not the programme is in a position to change certain context dimensions. These are then specified and it is stipulated by what means – activities – certain outputs and outcomes are to be generated in order to achieve the superordinate aim of the programme.

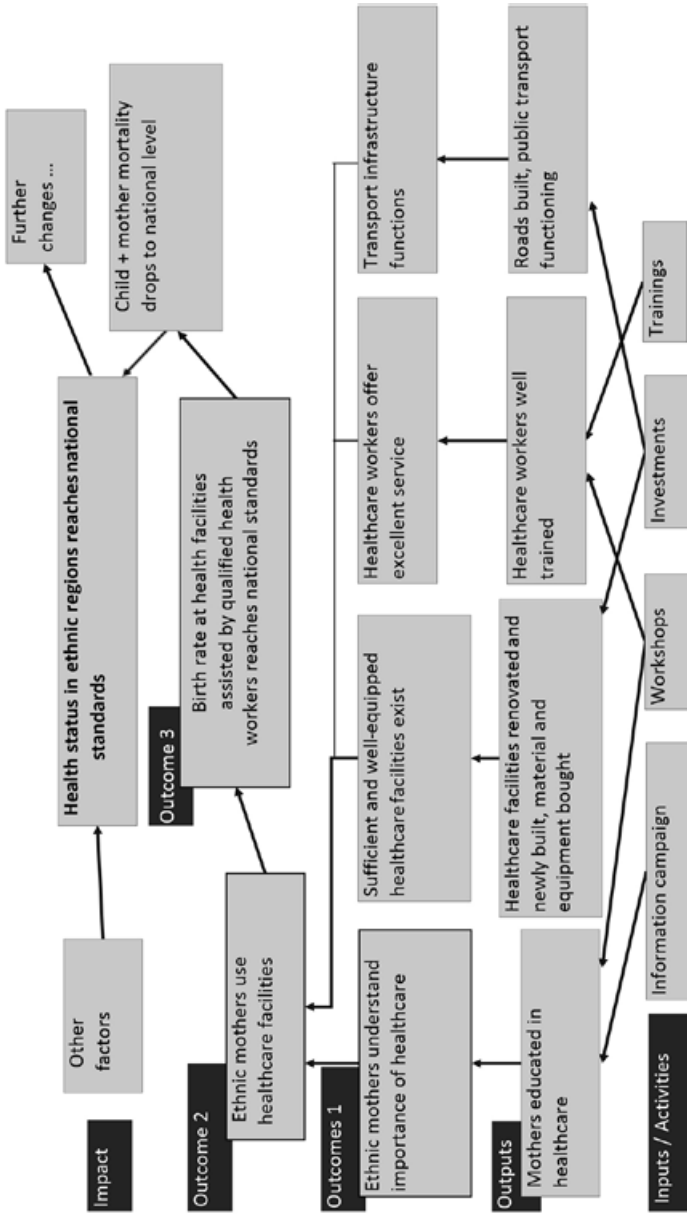


Figure 6.2 Theory of change – example: maternal health

SUMMARY

- ✓ Context is an amorphous term. What context is relevant for an evaluation can only ever be determined from situation to situation.
- ✓ Context is covered as an explanatory variable to a greater or lesser extent in the various different evaluation approaches, but they all agree ‘that context matters’.
- ✓ In the popular planning procedures too, context plays a role; this becomes especially clear in the formulation of programme theories, which are developed from the context.

As an *interim conclusion*, it can be said that in view of the diversity and amorphous character of context there is no such thing as an evaluation approach and a programme planning procedure that can cite all-embracing criteria for the selection of programme-relevant context factors. Instead, however, an evaluation approach should be in a position to deliver a solid foundation for the theoretically and methodologically essential incorporation of context factors in an evaluation. Stockmann’s CEval approach, which is structured systemically and comprises the micro, meso and macro levels of the context, meets these requirements in such a way that it is appropriate to refer to it briefly here.

6.1.3 The Context-related Nature of the CEval Approach

From a system theory point of view, the context consists of a large number of interdependent systems and subsystems. As shown in detail in Chapter 2, the CEval approach assumes that interventions (such as measures, projects, programmes etc.) are carried out by organisations in order to change context fields (subsystems) outside those organisations. Programmes etc. are temporary packages of measures for the initiation of processes of change. Which *organisational context dimensions* are to be investigated in an evaluation is a question derived from an organisation theory in the CEval approach. For this – in contrast to other evaluation approaches – there is a clear stipulation of organisation-related context factors which need to be incorporated, because they determine the performing strength of an organisation, on which, among other things, the possibility and the potential to implement a programme (or policy) and carry it through successfully depend. As well as organisation-strategic, structural and procedural elements (aims, organisational structure, technology), these include the financial resources of an organisation, and the employees with their qualifications, personal objectives, values, norms etc. – in other words the micro level. The organisational environment (context) is multiply interdependent with the *context that surrounds it (macro level)*, which can be divided down into subsystems. Which subsystems are relevant in an evaluation cannot be determined in a generally valid way in advance, but depends on the programme objectives, in other words on the *subsystems (contexts)* in which interventions are to take place and which other subsystems the subsystem in which the programme is located interacts with.

In the example programme from the healthcare sector cited here (cf. Figure 6.2), it would be necessary to examine whether or not the organisation commissioned to implement the

programme – the provider – is powerful enough to do the job or can be empowered by the programme. For that, the following organisational context conditions would need to exist:

- *Objectives*: the programme organisation actively backs the jointly developed theory of change and supports the agreed objectives of the programme, which are in harmony with the aims of the organisation.
- *Organisational structure*: the organisation has adequate functional (management) structures and (communication) processes which do not, at least, stand in the way of programme implementation.
- *Financial resources*: at first with financial support from a donor, the organisation has the financial means to build health centres, pay the employees appropriate salaries etc., and will also be in a position to finance the facilities thus established after the end of funding.
- *Technology*: the health centres have the technical equipment and application concepts that enable them to be run.
- *Employees*: the staff in the centres are obtaining or have obtained sufficient qualifications to be able to perform their duties.

Once these *organisational contextual conditions* exist, well-equipped health centres will exist, in which qualified employees work, who can advise the mothers as to the importance of healthcare for their children and carry out the necessary medical measures. In the example given here, other organisations – which can also be examined for their performance capability – are necessary too, e.g., to build roads and guarantee functioning transport facilities, so that the mothers and their children can actually reach the health centres.

The *external subsystems (contexts)* – i.e., those that lie outside the implementing organisation – which are important for the programme, ought to have been identified early on, in the programme planning phase and during the development of the theory of change, and can be investigated in an evaluation, or developed if none exist. In the example selected, these would include:

- the healthcare system (services provided, functionality, quality etc.);
- the economic system (possibility the families have to pay for health services);
- the cultural subsystem (standards, values, forms of behaviour such as would constitute an obstacle to going to a health centre: rejection of modern medicine, faith in traditional healers etc.);
- the transport infrastructure (accessibility of the health centres);
- the political system (willingness of the government to provide funds for the expansion of the healthcare system in ethnic regions in the long term etc.)

By means of this system or context analysis, the subsystems – context dimensions – which are important for the rectification of the state of affairs recognised as deficient can be determined, so that the activities initiated can lead to the planned outputs and the latter, in turn, can lead to the outcomes and aims of the programme that are considered necessary.

SUMMARY

- ✓ The CEval approach is a context-related approach which differentiates the context analytically according to internal (organisational) and external (environment-related) subsystems.
- ✓ The subsystems to be investigated in the organisational context are derived from the organisation theory.
- ✓ The relevant external – i.e., outside the implementing organisation – subsystems are identified during the development of the theory of change.

6.1.4 Evaluation as a Political Process

Not only measures, programmes or policies are interventions in a context, however that context is constituted, but also the evaluations themselves. Evaluations *per se* are geared to generating ‘evidences’, which are often accompanied by recommendations, on the basis of which – in the best case – people in positions of responsibility make decisions.

So evaluation is not just simply a scientific, methodological or technical activity, but a *political process* (Chelimsky 2008; Fitzpatrick, Sanders and Worthen 2012, p. 307; Rossi, Lipsey and Freeman 2019, p. 297). In the programme development phase, evaluation can have an influence on how social processes are perceived as situations identified as deficient, political problems etc., which theory of change is developed, which aims are defined for the programme and which target groups are selected. While the programme is being carried out, evaluations can influence the way in which measures are pursued, or even contribute to their being aborted. After conclusion of the implementation phase, evaluations say something about the effectiveness and sustainability of programmes and thus give legitimacy, for example, to the funding strategy of a financial backer or to government action.

Evaluation findings and recommendations thus automatically find themselves in an *area of conflict* consisting of various different perspectives, interests, expectations and aspirations of various different actors. This is about competing values, norms, political objectives, power and resources. That applies when projects or programmes are involved, and not less when it is a matter of complex policies and strategies which are intended to solve social problems. Intervening with its findings and recommendations in these social, political, cultural etc. contexts (subsystems), evaluation becomes part of them. By making object-related judgements and recommending that decisions be made on the basis of evidence-based conclusions, evaluation, unlike scientific research, is *highly political* – and, in turn, exposed to political attempts to influence it. Every evaluator needs to be aware of that.

Added to that is the fact that evaluation clients and evaluators live in *distinct ‘worlds with different logics’*. Regardless of whether the former are programme managers, administrative decision-makers or those in positions of political responsibility, the clocks tick at different speeds. For that reason, Rossi, Lipsey and Freeman (2019, p. 298) distinguish fittingly between ‘political time’ and ‘evaluation time’. Whilst some want quick results so that they can make decisions, the others are anxious to deliver methodologically ‘clean’ results, the generation of which sometimes takes more time (and sometimes more resources too) than the decision-makers are prepared to approve. If evaluation findings are to continue to be relevant

and significant for those responsible, it is often a question of finding a compromise between thoroughness of analysis and capitulation to the pressure of time and resources.

6.2 STAKEHOLDERS

6.2.1 Definition

In all the remarks made so far, the talk has been of protagonists, people affected and people involved, all of whom are covered by the term *stakeholders*.²

To specify more closely whom this term is supposed to designate, Greene (2005) suggests that distinctions should be made between four groups:

- people who have decision authority over the program, including other policy makers, funders, and advisory boards;
- people who have direct responsibility for the program, including program developers, administrators in the organization implementing the program, program managers, and direct service staff;
- people who are the intended beneficiaries of the program, their families, and their communities; and
- people disadvantaged by the program, as in lost funding opportunities.

Scriven (2007) adds yet more stakeholders to this list, and Fitzpatrick, Sanders and Worthen (2012, p. 289) also developed an extensive checklist for potential stakeholders, though it is best suited to programme evaluations.

Here, distinctions are made between the following groups of *stakeholders* (Brandt 2007, p. 166):

- *Clients*: those who commission the evaluation.
- *Evaluation funders*: those who finance the evaluation – these may be – but are not necessarily – identical to the clients.
- *Programme funders*: those who finance the programme or measure to be evaluated.
- *Programme managers, programme staff*: those who are directly responsible for the programme at strategic or operative level.
- *Target groups of the evaluation object*: persons to whom a programme, measure etc. is geared; programme participants.
- The *evaluation team* itself.
- *Other parties affected*: individuals or groups with no influence, or only a little influence on the way the evaluation is actually conducted, but are influenced by it or by the evaluation object itself (e.g., individuals who do not participate in a measure and may be at a disadvantage for that reason, programme competitors, individuals from the wider environment of a programme).
- *Citizens, parliamentarians, media* who may have an interest in the findings.

² Cf. Chapter 7 of this volume on stakeholder analysis and involvement procedures.

This and other detailed lists cannot be exhaustive, because the question of which stakeholders are relevant in an evaluation depends heavily on the evaluation purpose (cf. Chapter 2). In an evaluation, for example, in which it is a matter of further developing an ongoing programme focused on joint learning, and in which the evaluation results are not to be published, the broad-based participation of as many stakeholder groups as possible is more necessary than it is in an evaluation conducted for purposes of accountability legitimation.

As in the analysis of the context, performed to find out which factors, dimensions or sub-systems are relevant in an evaluation, the observation that there are no prefabricated, generally valid answers also applies to the *analysis of the stakeholder groups*;³ instead, the question is to be answered depending on the evaluation purpose and the context. The checklists referred to here can help to ensure that no important stakeholder groups are forgotten. Having said that, attention should be paid (as in the inclusion of various different context factors) to the fact that with an increasing number and diversity of the stakeholder groups, the complexity, the manageability, the time spent and the costs of an evaluation all increase too. For these reasons, prior to an evaluation, very considerable thought should be given to the question of who definitely needs to participate and, furthermore, exactly what that participation will involve (cf. Chapter 7.2 on stakeholder analysis).

Before we look at the various *possibilities* for stakeholder *participation*, another question needs to be cleared up: who has the right to speak for a stakeholder group?

In the case of the clients, the programme managers, this question can be answered quite quickly, but what is the situation with regard to the representatives of the target groups, the beneficiaries and other parties affected? If these groups are organised, they can delegate their representatives (for example, the mayor or village elder in a community, or the chairperson of an association, supervisory board etc.). But if that is not the case, it is far more difficult to identify stakeholders who are legitimised to speak for the whole group. Who is supposed to represent the mothers in the health project taken as an example above, or the impoverished rural population, the citizens in an urban district etc., if they do not have any institutionalised agencies?

Sometimes the proposal is made (cf. Lee 2004, p. 135ff.; Mertens 2004, p. 45ff.) that if no representatives can be found for the stakeholders, the evaluators should assume their role. However, that is not a very practical proposal. How are the evaluators supposed to know what the needs and points of view of such groups are?

To ensure the *representation* of selected stakeholders, at least, ‘advisory boards’ or reference groups are sometimes formed to accompany the evaluation process in an advisory capacity. But such bodies do not have advantages only. It is true that this is supposed to be a way of having external know-how flow into the evaluation, but there is of course, as a rule, an enormous amount of extra work associated with it in terms of communication. Moreover, it cannot always be assumed that such bodies have a positive attitude toward the evaluation. They can also be used by stakeholders to ensure that certain methods are not used or important target groups not surveyed, and/or that the processes of the evaluation are impeded in some other way.

³ In Chapter 7.2.1, a method for stakeholder analysis and stakeholder participation will be presented with the aid of ‘net mapping’. See Chapter 7.2.2, in particular Figure 7.1, on the management of stakeholder participation.

For example, in a large-scale evaluation of the sustainability of the environmental centres of the craft enterprises of a stakeholder group (which was represented on the advisory board), the group warned the evaluators emphatically not to carry out a telephone survey of the companies because the latter had no time to spare and better things to do. However, in view of the fact that the craft enterprises were the potential beneficiaries of the advisory services of the environmental centres, it was not possible to do without their assessments or their accounts of the experiences they had had. So the evaluators insisted on surveying them and were able to gain valuable insights by doing so. In another large-scale evaluation on the use of evaluation in organisations of German development cooperation, one of the organisations concerned actually went as far as to threaten legal action if the case study involving it were to be published.

The two cases described here are, of course, exceptions. In citing them, we aim to make it clear that evaluators are not automatically received with open arms by stakeholders and cannot automatically count on the latter's active cooperation when they announce that an evaluation will be taking place. Willingness to contribute to an evaluation and utilise its findings – in the author's 40 years of experience – is greatest if the evaluations are conducted for scientific or learning purposes; it is far less well developed if the evaluation is about accountability or legitimisation (whereby the results have to be published *per se*).⁴

SUMMARY

- ✓ Stakeholders are an essential part of context.
- ✓ In planning an evaluation, the groups of relevant stakeholders need to be identified.
- ✓ It is true that there are checklists for differentiating between groups of stakeholders but, as in the selection of relevant context elements, the selection of the stakeholders who will participate depends to a great extent on the purpose of the evaluation and the evaluation approach chosen.
- ✓ Advisory boards or reference groups can make a contribution to ensuring that selected stakeholders are represented on an ongoing basis, but they also cause extra communication work and can slow evaluation processes down.

6.2.2 The CEval Participative Evaluation Approach

Apart from the evaluation approaches which are participative *per se* and actively involve the various stakeholders in all the phases of an evaluation – right the way through to enabling them to conduct it independently themselves (Chapter 2.2) – there are numerous models in which stakeholder participation is not the dominant element, although it is planned to a greater or lesser extent.⁵ Almost all the authors agree that the participation of the stakeholders is important for the successful conducting of an evaluation and the utilisation of its findings (cf. Weiss 1998a, pp. 103–105; Patton 2008, p. 175; Fitzpatrick, Sanders and Worthen 2012, p. 286ff.).

⁴ Cf. Chapter 7.2.2 on the ethical dilemma between 'objective' credibility and the subjective usefulness of an evaluation.

⁵ On the question of the extent to which stakeholders should participate, cf. the remarks in Chapter 7.2.2, which are oriented to the 'ladder of participation' by Arnstein, 1969.

But if the stakeholders are going to be able to participate in an evaluation, they first need to be identified (cf. Chapter 7.2).

A series of advantages are associated with *stakeholder participation*, but there are risks too:

- First up, a procedure of this kind enables the various expectations of an evaluation, the various purposes to which the findings are going to be put, and the various perspectives of the evaluation object to be registered.
- Furthermore, stakeholder participation is associated with the assumption that apprehensions or fears, such as may be triggered by an evaluation, can be allayed.
- Moreover, it is assumed that by these means the acceptance of and support for an evaluation can be increased and a certain ‘ownership’ established, which in turn increases the chances that the evaluation findings will be implemented later on.
- By involving as many stakeholders as possible, interests can be balanced out better and the dominance of the client’s point of view can be relativised.
- Active stakeholder participation prevents the stakeholders from being seen as mere ‘data suppliers’; instead, they are perceived as emancipated partners with complementary interests.
- Since it can be assumed that the stakeholders have a broad-based knowledge of the evaluation context and the object to be evaluated – programme, policy etc. – valuable insider knowledge can be tapped and the validity of the results thus improved.
- Stakeholders can have a decisive influence on an evaluation. They can support it or impede it by passive resistance, for example by refusing to give interviews, refusing to be surveyed or to participate in other ways or, indeed, actively organising resistance to it.
- Intensive communication between the evaluators and the stakeholders facilitates a continual adaptation of the developed evaluation concept and the instruments deployed to changing contextual conditions during the course of the evaluation.
- It should not be forgotten that involving the stakeholders in the evaluation process can contribute to a situation in which they themselves gain new insights into the evaluation object and its context, understanding relationships better and developing new ideas.
- Lastly, participation in an evaluation can also lead to the general development of a positive attitude to evaluations or even, what is more, to the acquisition of evaluation knowledge such as can be used in future evaluations or programme work.

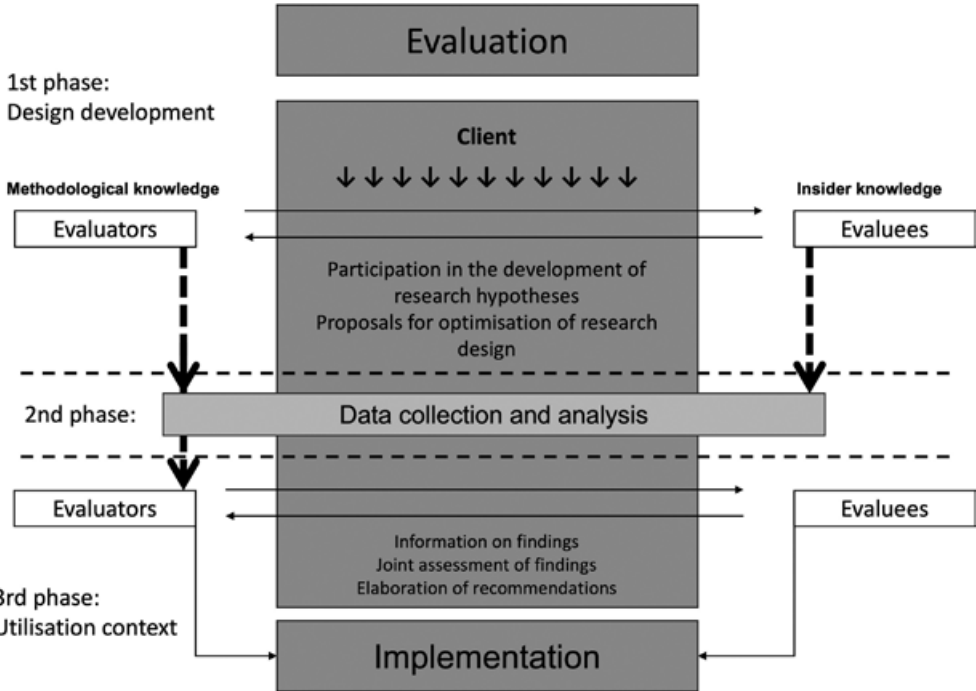
In order to benefit from the advantages of a participative procedure without renouncing the scientific aspirations of an evaluation, the CEval has developed a pragmatic approach which is based on the above-mentioned basic assumptions and has been tried and tested many times over in practice.

The procedure proposed here aims: (1) to conform to the scientific and professional standards, (2) to satisfy the information needs of the evaluators, (3) to showcase the various interests of the stakeholders as comprehensively as possible, (4) to ensure the scientificity of the evaluation, and (5) to guarantee its usefulness for the clients / stakeholders.

In the presentation that follows, it should be noted that this approach cannot simply be applied in a clichéd manner to any kind of external evaluation; instead, the degree or intensity with which stakeholders are actively involved in an evaluation depends on the evaluation purpose and – closely associated with it – the evaluation approach selected. Evaluators and stakeholders assume more or less complementary roles in this approach. Whilst the evaluators contribute their methodological knowledge, their moderation skills, their evaluation

experience and their professional competence, the stakeholders provide their subject-specific and actual knowledge of the situation, their precise knowledge of the programme to be implemented and the relevant context.

Figure 6.3 shows the respective functions of the evaluators and other stakeholders depending on the individual phases of the evaluation process. Putting it simply, the *ideal course* of an evaluation can be divided into three phases: design development, data collection and analysis, and utilisation.



Source: Stockmann 2008, p. 273.

Figure 6.3 The CEval participative evaluation approach

In summative evaluations these phases follow in a more or less chronological sequence. In formative evaluations they are to be understood as iterative loops, in other words each of the findings is ploughed back into the further development of the programme and the evaluation design, and the phases are repeated. (Cf. Chapter 5 on the planning and implementation of evaluations).

In the first phase of *design development*, it is above all the methodological knowledge of the evaluators that comes into play; in active collaboration with the stakeholders based on the latter’s knowledge of the subject and the field, they elaborate the evaluation questions, the evaluation criteria, the programme theory and the relevant design including the methods to be used (cf. Stockmann 2008, pp. 271–275 on this and in respect of the remarks that follow).

In the second phase, the evaluators *gather and analyse the data* on the basis of the design previously developed. Here, the stakeholders take on a special role as bearers of information. This way, the various different perspectives and assessments can be recorded in order to obtain an overall picture of the processes, structures and impacts to be evaluated, which is as comprehensible as possible and thus also 'realistic'. During the whole process, the evaluators keep the stakeholders informed as to the course and status of the project as a whole (for example, informally or via workshops with those involved), to ensure their integration in the process. Like data collection, data processing and analysis are also the responsibility of the evaluators.

In the *third and final phase*, the results thus processed can be discussed and assessed together with all those involved, with the evaluators taking on the role of moderators. Alternatively, an assessment already made by the evaluators can be presented to the stakeholders for discussion. Building on that, strategies for the further development of the evaluated programmes, measures etc. can be developed jointly, though actually implementing them is then the responsibility of the stakeholders and their organisations themselves.

In the approach developed here, *participative involvement* in an evaluation is concentrated primarily on the *design and utilisation phases*. The aims of an evaluation, the assessment criteria and, to a certain extent – provided that the scientificity of the design is not impaired – also the procedure can be established participatively and constitute the specifications for the evaluation. Gathering and analysing information by empirical-scientific methods, by contrast, is the job of the evaluators. The findings can of course be *assessed* together with the clients and the diverse stakeholders. The *utilisation* of the findings presented by an evaluation and their implementation in activities is, exclusively, the responsibility of the clients and the remaining stakeholders. Other than in quality management systems, the evaluator, particularly if recruited externally, is not part of the implementation process. The most he or she can do is to make recommendations, the realisation of which the clients and stakeholders are responsible for. The evaluator cannot influence these processes.

SUMMARY

- ✓ In the design phase, the active involvement of the stakeholders concentrates mainly on mapping out the aims of the evaluation and the theory of change, stipulating the assessment criteria and, to a lesser extent (provided that the scientificity of the design is not impaired), developing the methodological procedure.
- ✓ Those responsible for the collection and analysis of the data are, exclusively, the evaluators.
- ✓ In the utilisation phase, the stakeholders can be involved in the assessment of the results and the derivation of recommendations. The final realisation of the results is the responsibility of the decision-makers on the stakeholders' side.

6.3 ON THE RELATIONSHIP BETWEEN EVALUATORS AND STAKEHOLDERS IN THE EVALUATION PROCESS

Now that the term ‘stakeholder’ has been specified more precisely, the importance of the stakeholders and the opportunities and risks associated with that importance mapped out and – building on that – a participative evaluation approach presented, this chapter is about the *role of the evaluator in the evaluation process* and his or her relationship with the stakeholders. A look is taken at the various different requirements relating to evaluators working internally or externally and areas of conflict that may crop up, to ensure that the evaluation does not just follow an optimum course, but also generates a high level of benefit.

Accordingly, there needs to be a distinction between whether an evaluation is conducted internally or externally (cf. Chapter 2.3.1). In *internal* evaluations, evaluators have the *advantage* of being very familiar with the programmes being carried out and the aims, problems etc. associated with them; in other words, they have concrete situational knowledge and can build on existing communication structures and begin swiftly with the evaluation, because no unwieldy invitation to tender or elaborate assignment procedure is necessary (cf. Stockmann 2008, p. 263; Volkov 2011, p. 26ff.; Fitzpatrick 2012, p. 27f.; Böttcher 2020, p. 1552f.; Owen 2021, p. 142ff.; and others). At the same time, because of the organisational propinquity to the decision-makers and programme managers, recommendations can be made more quickly and implemented promptly.

However, that propinquity can turn into a *disadvantage* if the evaluators – on account of the organisational structures and constraints to which they are bound as members of the organisation – fail to maintain the necessary objectivity to the content of the programme and, in particular, also to those acting in it. Thus, on the one hand, there is a danger that internal evaluators may lose the necessary critical distance to the evaluation object. This may (under certain circumstances) lead to situations in which they are, for example, not open enough to alternative explanations, models or procedures. On the other, there is a risk that, on account of their dependence on superiors, internal evaluators may shrink from making critical assessments of the programme and pay too much attention to the latter’s points of view, as a result of which other, less influential groups within the organisation are disadvantaged (cf. Stockmann 2008, p. 263; Strobl, Lobermeier and Heitmeyer 2012, pp. 158–160).

In self-evaluations, these challenges may intensify, because the evaluators are confronted with having to play two different roles at the same time: their task as evaluators and the implementation management of the programme to be evaluated.

The advantages and disadvantages that accompany the framework conditions for *external* evaluators are almost the complete opposite of those of internal evaluations (cf. Stockmann 2008, p. 263; Fitzpatrick 2011, p. 28; Owen 2021, p. 132ff.). As a rule, for example, external evaluators have more distance to the evaluation object and also to those participating in the programme and other stakeholders, which ought to give them a more objective view and a greater degree of independence. By contrast, the access they have to the evaluation environment can be more difficult to gain than the access available to internal evaluators. Internal evaluators are thus familiar with the problems and the language of the various different actors and may already have built up an intensive relationship of trust with them while implementing the programme. By contrast, external evaluators come into the field from outside and thus first have to gain access to the protagonists, adjust to their idiosyncrasies and points of view and learn how to deal with them. Having said all that, even if there are different situational require-

ments for internal and external evaluators, the same professional standards apply to both, as defined, for example, by the DeGEval.

Other important factors that influence the role of the evaluator and his or her relationships with the stakeholders depend on *the evaluation assignment* and the *evaluation purpose* associated with it, the *time phase* of the programme, the *analysis perspective* etc. (cf. Chapter 2). The evaluators need to adapt to these factors accordingly (cf. Weiss 1998a, Chapter 5; Patton 2008, p. 213; Kunding 2010, p. 350ff.; Fitzpatrick 2011, p. 18; Böttcher 2020, p. 1558) In formative evaluations, for example – the results of which are intended to serve the further development of a programme – continuous, strongly interactive dealings with the stakeholders are useful in strengthening their willingness to participate and their motivation to implement evaluation recommendations. So for a purpose of that kind, it is participatively oriented evaluation approaches which tend to be deployed, giving the stakeholders an active role in the planning and implementation of the evaluation. In summative evaluations, which serve above all purposes of accountability or legitimation, there is as a rule no such wide-ranging stakeholder involvement. In these cases, management-oriented or programme-theory-oriented approaches tend to be deployed, which are accompanied by more aloof role behaviour on the part of the evaluators with the aim of demonstrating a higher degree of independence, which plays an important role in the credibility of the results.

Regardless of which evaluation approach is used and regardless of whether an evaluation is conducted internally or externally, the *communication* between evaluators and stakeholders always plays a decisive role. The underlying relationship with the protagonists involved in an evaluation and those affected by it and the social interactions among them are central influencing factors in an evaluation. As shown at the beginning, the various different interests, moral values and cultural views among the individual groups of stakeholders have an important role to play. But that also applies to the *evaluator*. Far from being a ‘*neuter*’ hovering weightlessly above the whole, the evaluator too is integrated in a system of values and norms. Moreover, he or she is part of a complex network of relationships, which consists of the various different actors in the evaluation context, and to whom he or she may be obliged from a moral – ethical – personal, financial, organisational or other point of view. An evaluator must be aware of this fact and must endeavour, via continual *self-reflection*, to monitor his or her own behaviour, and not to take sides in favour of any particular group of stakeholders or take his or her own moral values as a yardstick for the assessments made (cf. Kunding 2010, p. 354f.; Fitzpatrick, Sanders and Worthen 2012, pp. 79–96, 97–101).

Possible examples of how personal attitudes and value positions of the evaluator can cause interference are *ethnocentrist attitudes* or the danger of ‘*going native*’. Ethnocentrist distortions may come about (unintentionally) if researchers move in unfamiliar cultural contexts and apply their theories, methods and instruments, which were developed for the analysis of western societies, without thinking and without subjecting them to sufficient cultural adaptation. Even in the formulation of the problem to be investigated and the presumed relationships, such critical-rational research logic can run into comprehension problems in unfamiliar cultural contexts. ‘Going native’ is a phrase used to describe the tendency to adopt the views of individual actors or stakeholders. If they do that, the evaluators lose their distance to the object being investigated and their independence in respect of certain interests. The longer an evaluator remains in that milieu and the more intensive his exchange of information with the stakeholders, the greater this danger becomes.

The evaluator's *relationship with the client* can be particularly problematic, being both financial and organisational. However, there is a distinction to be made here between internal and external evaluation. In an internal evaluation there are no 'classical' clients as there are in an external evaluation. Nevertheless, dependences do exist. The management of an organisation, for example, can decide which programmes are to be evaluated, how much money is to be made available, how the evaluation findings are to be handled, what degree of appreciation is to be shown to the internal evaluators, and how the work they do will affect their further careers etc.

In external evaluations, it is assumed that the evaluators have a greater degree of independence, given that they are not employees of the organisation that engaged them and thus cannot be sanctioned directly. However, that does not change anything in respect of the fact that it is the client who remunerates them financially. In particular, dependences may arise in that respect if an evaluator works for a given client very frequently. The more heavily his economic situation depends on a follow-up assignment, the greater the risk that the assessment of the findings may be biased, because: 'The possibility of future evaluation contracts or consulting depends on how well the client likes the most recent evaluation completed by the evaluator' (Fitzpatrick, Sanders and Worthen 2012, p. 102). There are, moreover, organisational dependences. The client controls access to resources, logistics, stakeholders who could be involved or surveyed etc. and can thus influence the evaluation results.

6.4 EVALUATORS UNDER PRESSURE

6.4.1 Attempts to Influence the Evaluators and Patterns of Criticism

For evaluations to be able to generate reliable, valid and credible results, *scientific independence* is necessary. For that reason, it is called for in the standards, guidelines and codes of national and transnational evaluation societies and organisations all round the world. For example, the DeGEval fairness standards (which will be looked at more closely later on) state rather clumsily that 'it should be clear throughout the evaluation process and from the evaluation reports that the evaluators are impartial' (<https://www.degeval.org/home/>). The OECD's DAC Evaluation Quality Standards explicitly feature two paragraphs on the independence of evaluators vis-à-vis those involved:

- In the evaluation report, information about the degree to which those who make assessments are independent of the policies, measures and administration of the client, those who implement the evaluation object and those who benefit from it is provided. Possible conflicts of interest should be addressed openly and honestly.
- Free and open assessment process. The evaluation team is in a position to work freely and without interference. The team should receive assurances as to cooperation and access to all the relevant information. The evaluation report should mention any hindrances that may have had an effect on the evaluation process (<https://www.oecd.org/dac/evaluation/qualitatsstandardsfurdieentwicklungsevaluierungsentwicklungsausschusses.htm>).

These standards are important, because a series of studies have shown 'that in all countries, evaluators are confronted with attempts by stakeholders to influence evaluators to misrepresent their findings' (Stockmann, Meyer and Schenke 2011; Morris and Clark 2013; LSE

GV314 Group 2014; Pleger and Sager 2016; Pleger et al. 2017, p. 316). These *attempts to influence the evaluators* relate to all the phases of the evaluation process and are dominated by the contractors.

	Studies		
	Open-Ended Question	Closed Question	
	Survey, USA (n = 340)	Survey, Germany (n = 121–123)	Survey, Switzerland (n = 61–63)
Content of Misrepresentation Request			
Present findings more positively	38% (n = 130)	78% (n = 96)	76% (n = 48)
Present findings more negatively	Category not offered	30% (n = 37)	27% (n = 17)
Omit or downplay negative findings	25% (n = 85)	Category not offered	Category not offered
Change language—neither positive nor negative	13% (n = 43)	80% (n = 98)	87% (n = 54)
Use inappropriate/different methodology or statistical procedures	6% (n = 21)	72% (n = 87)	63% (n = 39)
Draw different conclusions	6% (n = 20)	53% (n = 65)	56% (n = 35)
Show inappropriate concern for implications of results	4% (n = 12)	Category not offered	Category not offered
Use invalid/old/different data	3% (n = 11)	57% (n = 69)	45% (n = 28)
Distortion of content	Category not offered	55% (n = 68)	55% (n = 34)
Results were determined in advance by the stakeholder	Category not offered	Category not offered	41% (n = 25)
Other	5% (n = 18)	Category not offered	30% (n = 16)

Source: Pleger et al. 2017, p. 322.

Figure 6.4 Changing requirements of stakeholders

Figure 6.4 provides an inter-country overview of the types of attempt to influence evaluators. As can be seen, these are very wide-ranging, and the results obtained from Switzerland and Germany differ markedly from those obtained from the USA. Clearly, the problem of being influenced by clients in the former two countries is more pronounced than it is in the USA. Requests to ‘present findings more positively’, however, are made particularly often in all three countries. In Germany 78%, in Switzerland 76% and in the USA only 38% of the evaluators surveyed said that they had been exposed to such an attempt (to influence them) on at least one occasion. In addition, a high percentage of those surveyed from Switzerland and Germany said that they had been requested to use different language or to deploy methodologies or analysis methods that were inappropriate.⁶

The study of the independence of evaluations in Germany (Stockmann, Meyer and Schenke 2011) shows – astonishingly – that a relatively large proportion of the DeGEval members surveyed said they had yielded to various attempts to influence them on at least one occasion. More than 90% said they had obeyed a request to reword individual sentences in presentations

⁶ For an explanation of the differences between countries cf. Pleger et al. 2017.

of their findings. Given that such a request does not necessarily constitute an intrusion into the independence of an evaluator – in the opinion of the evaluators surveyed, at least – this can be classed as unobjectionable. By contrast, the fact some 60% of the external evaluators said they had catered to the wish of a client to present findings in a more positive light (Stockmann, Meyer and Schenke 2011, p. 66) should provide much more food for thought.

If we look at the frequency of attempts to influence the evaluators by the fields in which they were working, we see that at 40%, the field of development cooperation was well above the average (26%) (Stockmann, Meyer and Schenke 2011, p. 63). This field distinguishes itself by having only a few clients, many providers, and thus a high pressure of competition, and that may be the reason why it is particularly susceptible to such attempts to influence evaluators.

As reported by those surveyed, attempts of this kind are seldom made in a crude fashion, threatening the evaluators perhaps by saying that they may not receive a follow-up assignment, or that part of the agreed fee may actually be withheld, but usually more subtly, for example by expecting them to show some understanding for the consequences of critical statements for those affected, for the organisation or in terms of public perception.

If these attempts to influence the evaluators are systematised, we can – regardless of the number of mentions – differentiate between the following *interventions on the part of clients* in the evaluation process.

- (1) In the planning phase:
 - the client defines desirable or undesirable findings early on, while the evaluation is still being prepared;
 - attempts to mask out certain questions for ‘political’ reasons from the very outset, although in fact they are relevant;
 - requests not to include certain groups of stakeholders in the evaluation, so that their interests and points of view are not recorded or taken into account.
- (2) In the implementation phase:
 - attempts not to use certain methods for data collection or data analysis;
 - the wish to name informants who had been assured of anonymity by the evaluators, or to pass on confidential information.
- (3) In the utilisation phase:
 - attempts to put the evaluators under pressure to present findings differently (more positively) or suppress them completely, or to manipulate them in some other way;
 - use of evaluation findings by the clients to impose sanctions on individual actors (stakeholders) including their own employees;
 - conscious attempts to conceal the findings of an evaluation from superiors or the general public, right the way through to non-compliance with agreements to publish evaluation reports in an unabridged and uncensored form;
 - conscious abuse or misinterpretation of evaluation findings by stakeholders.

Such attempts to influence the evaluators, and attempts of other kinds caused by clients, are *potential causes of conflict* in evaluations. But other groups of stakeholders too, such as those responsible for the programme, i.e., programme managers, programme staff, participants and persons who have been deliberately or accidentally excluded from a programme, can disrupt the planned course of an evaluation.

Given that the tasks of evaluations include revealing deficiencies and aberrations and rendering them transparent, even if that means casting doubt on the strategies and political positions of the stakeholders, in particular those of the client, or even if all the stakeholders are integrated in the best possible way, it cannot be assumed that everyone will always be thrilled about the evaluation results:

This means that sponsors of evaluation and other stakeholders may turn on the evaluator and harshly criticise the evaluation if the results contradict the policies and perspectives they advocate. Thus, even those evaluators who do a superb job of working with stakeholders and incorporating their views and concerns in the evaluation plan should not expect to be acclaimed as heroes when the results are in. The multiplicity of stakeholder perspectives makes it likely that no matter how the results come out, someone will be unhappy. (Rossi, Lipsey and Freeman 2004, pp. 43, 374)

Because evaluations take place in a *political environment*, it sometimes happens that stakeholders react dismissively to results which contradict their own positions and expectations. Attempts may even be made to discredit the evaluation and/or those who have conducted it. So evaluators should not be all that surprised if their study or indeed they themselves end up in the crossfire of *criticism*.

Many years of evaluation practice enable the following typical types of criticism to be identified:⁷

- (1) *So what's new?* Sometimes, the clients of the evaluation, evaluatees or other stakeholders claim that the findings of the evaluation were already known to everyone before the evaluation and thus come as no surprise. Often enough, it is indeed true that those affected are aware of deficiencies and problems or have at least developed a feel for them. Having said that, the task of the evaluation is to deliver empirically founded verifications and robust findings. In such a case, however, regardless of that, the question of why, in spite of their knowledge of the existing problems, those responsible have not taken action, and why the 'familiar' deficiencies were not remedied prior to the evaluation, does need to be cleared up.

For the rest, in an evaluation, as throughout social science, findings which are counterintuitive, i.e., findings which do not correspond to the general expectations, can be seen to attract the most attention. Yet empirically evidence-based insights which are in harmony with the mainstream of implicit or explicit suppositions and assumptions are no less important.

- (2) *Methodological deficiencies:* A particularly popular option for casting doubt on the results of an evaluation consists of criticising the design and methods used. Given that many different approaches to an investigation problem are indeed possible, only the selection of an appropriate design and practicable methods offer protection against unjustified criticism. For these reasons an evaluator must prove convincingly that his methodological procedure duly corresponds to the 'state of the art'.

A problem may arise if the client does not make the necessary funds available to finance an evaluation design which is genuinely appropriate to the tasks at hand. Often, the art of evaluation consists of just that kind of challenge: obtaining an optimum amount of robust information with a minimum of means. 'Second-best' approaches

⁷ Text taken largely from Stockmann 2008, pp. 268–270.

often have to be accepted, since the evaluation cannot be conducted in any other way for lack of resources in terms of money and/or time. In spite of that, there are clients who only provide a modest amount of funds for an evaluation and generously accept a 'second-best' solution, but then go and criticise the methodological shortcomings at the end of the study. In order to be able to prove that, during the preparation for the evaluation, attention was drawn to the methodological difficulties and to the consequences for the quality of the evaluation associated with a 'second-best' solution, evaluators are recommended to document all the steps and to adopt all the statements made en route together with the client.

When all is said and done, however, *responsibility for the quality* of an evaluation lies with the evaluators. For that reason, if they see that the prerequisites for conducting an evaluation such as would be appropriate to the task do not exist and that the standards for evaluation cannot be adhered to, they should turn the job down.

The interview partners should be selected with the greatest circumspection. Given that representative (random) selections are often not possible, evaluators should proceed very carefully here. Care should be taken to ensure that, if possible, all the relevant points of view and interests are represented. Furthermore, evaluators are recommended to agree the selection with the stakeholders (or at least with the client). Otherwise, they may be accused of having interviewed the 'wrong' people and thus arriving at 'incorrect' or distorted findings. If – the accusers may say – the 'right' people had been asked, the assessment would have been quite different, i.e., more positive.

- (3) *What must not be true cannot be true*: Sometimes, findings are simply denied. Obviously, if they are facts and circumstances that can be proved beyond doubt by means of data, the situation can be cleared up quickly. If it is a question of statements of opinion (for example, satisfaction with various aspects of a programme), evidence should be provided via a statistically adequate number of interviewees. The data then 'speak for themselves'. If, however, it is a matter of interpretations on the part of the evaluators, care should be taken to stick strictly to a logical chain of reasoning. The more statements are avoided that cannot be evidenced with existing data, the smaller the target offered by an evaluation. And finally, no evaluation should get involved in speculation.

Especially in the case of very complex evaluands, it is often not possible – in spite of the very greatest diligence – to preclude the possibility that factual errors may be made. If such errors are complained about by those responsible for or affected by a programme who, as a rule, have far more situational knowledge than the evaluators, they are of course – following detailed investigation – to be corrected. But the position is different when it comes to reasoned assessments. Evaluators not only have the right, but also a professional duty to stick to judgements which are sufficiently evidenced with facts, and to resist all possible attempts to influence them.

- (4) *Painstaking search for errors*: In presentations of the findings such as are summed up in the evaluation report, criticism by the client or the evaluatees can sometimes be seen to exhaust itself in innumerable details. Given the latter's superior concrete situational knowledge, it is almost always possible to discover errors in the presentation, even if they are only marginal. Even grammatical errors or incorrect punctuation in the final report can provide fuel for debates. In cases like that, care should be taken to ensure that the main statements and insights from a study do not get shoved into the background.

The fact is that there may be method in such a manoeuvre, for example an attempt not to have to come to terms with any unwelcome statements the study may be making.

- (5) *Implementation deficiencies*: Not all evaluations are well planned. If it happens that the client does not deliver the promised support, for example failing to provide the addresses which are necessary for a survey, if designated interview partners are never available, or if processes and decisions keep on being postponed, these problems must be documented by the evaluator in detail. That is the only way in which evaluators can defend themselves against subsequent accusations – for example, that the number of interviewees is too low, or that the report has not been completed within the deadline agreed. And it hardly needs to be said that the evaluator must draw the attention of the client to such problems and – as far as possible – assist him in solving them.

At this point we should make it quite clear that attacks of this kind may occur, but that they are by no means the rule. We do not wish to give the impression that criticism of evaluation studies or their authors is always without reason and attributable to the inability of clients, evaluatees or other stakeholders to withstand criticism or their lack of willingness to learn. Certainly not! It goes without saying that evaluations quite often give occasion for justified criticism. Thus not every critical utterance is to be classified as an attack; instead, a review should be undertaken in all cases as to whether or not the criticism is justified. Neither does every proposal for an amendment constitute an attempt to influence the evaluators (cf. Perrin 2019, p. 355).⁸

As a matter of basic principle, it may be assumed that the clients of an evaluation, who are, when all is said and done, the ones who provide the financial, organisational and logistical resources for its implementation, have a genuine interest in the findings, and wish to put negative findings to constructive use for the further development of their programmes, policies etc. This applies particularly to organisations in which the practice of dealing constructively with criticism already exists, in which quality discussions are held openly and errors are allowed to be referred to so that they can be learned from. In organisations in which such a ‘*culture of learning*’ exists, evaluation findings are more likely to be accepted and recommendations – assuming a professional evaluation – more likely to be implemented than they are in organisations in which that is not the case.

SUMMARY

- ✓ The independence of the evaluator is a valuable asset that should be protected.
- ✓ Attempts to influence the evaluators are mostly made by clients, since they, as the financing entity, have a certain potential for exerting pressure.
- ✓ Attempts to influence the evaluators are usually subtle, come in many different guises and can take place in any phase of the evaluation process.
- ✓ Not all criticism is justified, but on the other hand not every critical utterance and not every proposal for alteration constitutes an attack on the independence and integrity of the evaluator.

⁸ It follows that the accusation made by Perrin (2019, p. 354), that the authors (Pleger et al. 2017) ‘seem to suggest that any pressure on evaluators to change their reports may be inappropriate’ is nonsense.

- ✓ All points of criticism brought forward by stakeholders in the evaluation process or regarding the evaluation report are to be investigated very thoroughly by the evaluator with regard to their validity.

The best protection against criticism is *professional work*. That includes:

- (1) an *active communication strategy*, which entails a transparent flow of information from the very beginning, and communication among the protagonists, and – depending on the purpose of the evaluation and the approach deployed – gives them the right to participate (cf. Chapter 11);
- (2) *far-sighted planning* as regards the resources available in terms of time and finances so that the individual operations of the evaluation can be implemented (cf. Chapter 5);
- (3) the development of a *programme theory* which is appropriate to the context and shared by as many of the stakeholders as possible (cf. Chapter 6.1.2);
- (4) use of an *evaluation design* which is appropriate to the evaluation purpose and the research questions, and a suitable *data collection and data analysis procedure* (cf. Chapters 7–10);
- (5) adherence to professional, academic and ethical *standards* so as to ensure usefulness and feasibility and fair dealings with one another (cf. Chapter 2.3.3);
- (6) taking into account the various different interests and points of view of the *stakeholders* when planning and conducting the evaluation and interpreting its results, so that the evaluation can generate the greatest possible benefit (cf. Chapter 6.6); and
- (7) that requires the *results* to be *presented* in a way which is not only scientifically well founded, but also worded appropriately for the audience in question, and the recommendations made to have been derived logically from the findings and conclusions arrived at (cf. Chapter 11).

6.4.2 Fairness Standards and Ethical Principles

For the processes of social interaction between evaluators, clients and other stakeholders to be able to run as fairly and respectfully as possible, the various evaluation societies have laid down so-called fairness standards as a kind of guide. The DeGEval, for example, has agreed the following standards (<https://www.degeval.org/home/>):

FAIRNESS

F1 – Formal agreements

The rights and obligations of the parties involved in an evaluation – what should and may be done, how, by whom and when – should be set out in writing.

F2 – Protection of individual rights

Evaluations should be planned and conducted in such a way that the rights, security and dignity of those concerned are protected.

F3 – Comprehensive and fair investigation

Evaluations should investigate and depict the strengths and weaknesses of the evaluation object as fairly and comprehensively as possible.

F4 – Impartial implementation and reporting

The evaluation should consider the various different points of view of the stakeholders as regards the object and results of the evaluation. The whole evaluation process and the evaluation reports should reflect the impartiality of the evaluators.

F5 – Disclosure of findings and reports

Evaluation results and reports should, as far as possible, be made accessible to all those involved and all those affected.

Other societies have adopted similar standards. Here is a selection:

- Joint Committee on Standards for Educational Evaluation: The Program Evaluation Standards (2010) (<https://evaluationstandards.org/program/>).
- African Evaluation Association (AFREA): The African Evaluation Guidelines (2020) (<https://afrea.org/wp-content/uploads/2020/03/AEG-29-February-2020-FINAL-DRAFT-for-consultation.pdf>).
- Latin American and Caribbean Network of Monitoring, Evaluation and Systematisation (ReLAC): Evaluation Standards for Latin America and the Caribbean (2021): (<https://www.relac.net>).
- OECD/DAC Principles for Evaluation of Development Assistance (2000) (<https://www.oecd.org/dac/evaluation/dacprinciplesforevaluationdevelopment-assistance.htm>).
- OECD/DAC Quality Standards for Development Evaluation (2010) (<https://www.oecd.org/development/evaluation/qualitystandards.pdf>).
- United Nations Evaluation Group: Standards for Evaluation in the UN System (2005 [updated 2016: norms and standards]) (<http://www.uneval.org>).

Whilst these standards refer to the evaluation process and define criteria to which an evaluation should adhere, there are also standards which make direct specifications on the way the evaluators should behave. The most influential code is probably the *Guiding Principles for Evaluators*, published by the American Evaluation Association in 1994. These feature five guiding principles:

- (1) *Systematic inquiry*: Evaluators conduct data-based inquiries that are thorough, methodical and contextually relevant.
- (2) *Competence*: Evaluators provide skilled professional services to stakeholders.
- (3) *Integrity and honesty*: Evaluators behave with honesty and transparency in order to ensure the integrity of the evaluation.
- (4) *Respect for people*: Evaluators honour the dignity, well-being and self-worth of individuals and acknowledge the influence of culture within and across groups.
- (5) *Common good and equity*: Evaluators strive to contribute to the common good and advancement of an equitable and just society.

Such standards / guidelines can help to mitigate conflicts that arise either preventively or during the evaluation, though of course conflicts cannot be prevented altogether.

6.5 COMPETENCE REQUIREMENTS FOR EVALUATORS

6.5.1 Multi-faceted Requirement Profile

The previous remarks have shown that evaluators are confronted with a wide range of challenges and great expectations. To fulfil those expectations, they need extensive, wide-ranging *competences*.

EXCURSION: COMPETENCE OR QUALIFICATION?

Even if in our everyday understanding of them the words ‘competence’ and ‘qualification’ are often used synonymously, there are differences between them: qualifications denote the knowledge, abilities and skills that enable a person to pursue a (work-related) activity. They can be evidenced by final examinations and documented by certificates. Accordingly, the term qualification is used when demand-oriented characteristics are meant. For example, if the aim is to express the fact that a person has the qualifications to pursue a certain activity or profession, in other words that he can meet the actual requirements.

The term competence, by contrast, is broader, and directed at a person’s overall ability to act. In other words, competences include qualifications, but they also include the ability to cope with new, dynamic situations. Competences thus refer to general dispositions people have to cope with situation-specific requirements for action. As long ago as 1974, the German Education Council defined vocational competences, as against qualifications, as abilities, skills, bodies of knowledge and attitudes that enable the individual to act in a completely professional and social way (www.bibb.de/de/8570/php; Moore and Theunissen 1994; Häcker and Stapf 1996; Schaeper 2005a, 2005b; Friebe 2006).

With the change from an industrial society to a society of knowledge, the requirements relating to competence, which can be divided into professional, social and self-competence, have been displacing the qualification requirements which used to be more dominant, because in a dynamic and complex environment it is above all ability to act that is called for.

This change of concept, however, is also somewhat blurred in notional terms, since competences cannot be defined or measured as precisely as qualifications. Nevertheless, the term competence is used in the section that follows because it fits in ‘better’ with the task profile of the evaluator.

Evaluation competence comprises all the knowledge, skills and proficiencies (qualifications) and the social and self-competence to be able to implement professional, high-quality evaluations which are flexibly adapted to the respective situational context.

Given that an evaluation is not merely supposed to meet scientific and academic criteria, but also to be conductible in the framework of predetermined resources paying heed to the fairness

standards cited above, and useful to diverse stakeholders as well, specialised scientific knowledge and methodological skills are simply not enough. To be able to work in an evaluation team and do justice to the stakeholders with their different points of view and interests, *social competences* are called for too, such as the ability to work as part of a team and the ability to communicate: all the more so if the evaluation is formative or participative and one in which group processes are to be moderated, conflicts of interest balanced, common future perspectives developed and potentials for improvement tapped.

As we have said, evaluators act in a nexus of power and interests, are part of a political process and often work in a culture that is unfamiliar to them, so that they should also, and last but not least, have skills involving *self-reflection* so that they can determine their own position.

Organisational competences are vital if there is to be a guarantee that an evaluation will be planned, managed and conducted professionally. For it to succeed, of course, *specific evaluation competences* are also required, for example to be able to develop a workable set of evaluation questions; to penetrate the causal mechanics of the evaluation object with the aid of the programme theory; to use an evaluation approach, an appropriate design and a method concept which are adequate for the problem at hand; to guarantee a professional analysis and interpretation; and, at the end of the day, to submit an evaluation report that will be useful to the stakeholders. All this presupposes a fundamental understanding of the connections between evaluation object, questions, concepts and methods, which calls for broad-based fundamental knowledge in social-science and evaluation-related terms.

Context-related competences continue to be important: for the appropriate evaluation of an object, subject-specific, sectoral and general contextual knowledge is required. This includes content subsumed under the term subject-specific competence or field competence, such as basic knowledge of the subject at hand and knowledge relating to the existing structures and framework conditions in a given area of practice. On the other hand, it also includes content that – generally, and independently of specific areas of application – covers competences required for the analysis of organisational and political-administrative relationships. If one is out and about in unfamiliar cultural contexts, knowledge of the religious and cultural attitudes of the people and the way they behave is necessary, as are appropriate linguistic skills.⁹

If evaluators do not have all the necessary competences, it is a good idea to form *evaluation teams* which can compensate for that. It is, for example, usual to make up for lacking field, cultural and linguistic competences with (native) experts in these areas.

In order to strengthen the specific evaluation competence of native evaluators, appropriate *capacity building* measures have been available for decades now, and in recent years, through the foundation of the Global Evaluation Initiative (GEI) (<https://www.global-evaluation-initiative.org>), attempts have been made to package them. For 20 years, the International Program for Development Evaluation Training (IPDET), which has been running since 2018 at the University of Bern in cooperation with the CEval in Saarbrücken, has also been contributing to this. The IPDET offers a basic one-week course, consolidation topics that build on it and special target-group-related courses (<https://ipdet.org>).

⁹ On competence requirements for evaluators cf. Scriven 1996; Stevahn et al. 2005; Schwandt 2008; Patton 2008, p. 199ff.; Cooksy and Mark 2012; Fitzpatrick, Sanders and Worthen 2012, p. 231ff.; King and Stevahn 2015.

6.5.2 Competence Catalogues and Teaching / Learning Goals

As described at the beginning, one of the central problems is that ‘evaluator’ is neither an academic nor a vocational title. Evaluators are, as a rule, not licensed or certified, nor do they have an academic degree in evaluation.

Instead, they come from various different disciplines: in the USA they will mostly have studied educational science or psychology, and in Europe more likely political science, sociology or economics. Over time they will have learned evaluation ‘on the job’ (cf. Fitzpatrick 2012, p. 17). Rossi, Lipsey and Henry (2019, p. 300) thus describe evaluators ‘as a collection of individuals sharing a common label, who are not formally organized, and who may have little in common with one another in terms of the range of activities they undertake or their approaches to evaluation, competencies, organisations within which they work, and perspectives’.

Basically, in fact, anyone – regardless of his or her background in terms of education or training – can conduct an evaluation, and that of course has led to a situation in which evaluations do not always attain the quality standards which are necessary from a professional point of view.

To counteract this, there have been two developments in recent years, which run parallel to each other but are relatively unconnected. One is that national evaluation societies such as the Canadian Evaluation Society, the United Kingdom Evaluation Society, the American Evaluation Association, and transnational societies such as the European Evaluation Society and the International Development Evaluation Association, but also professional networks such as the United Nations Evaluation Group and the Evaluation Cooperation Group (of multilateral development banks) have developed so-called ‘*competency frameworks*’.

On the basis of a comparative analysis of some of these frameworks (UNEG 2016; AEA 2018; IEG 2019) the following *competence fields* for evaluation practitioners have been identified:

- evaluation expertise (including a broad understanding of the purposes, institutionalisation, processes and approaches of evaluation);
- methodological expertise;
- management expertise;
- communication and interpersonal skills;
- integrity and ethics;
- institutional expertise;
- substantive expertise; and
- contextual expertise (Vaessen, Joppert and Stockmann 2022).

Some societies have not merely developed competence catalogues, but also offer a kind of *certification* with the title ‘credential evaluator’ (<https://evaluationcanada.ca>).¹⁰

The other trend to be seen in terms of pushing ahead with the professionalisation of evaluation is the *establishment of academic study courses* which lead to a master’s degree, a diploma or even a doctorate degree. As the CEval’s Evaluation Globe Project has shown, amenities of this kind are offered at many universities in Europe and North America, but hardly at all

¹⁰ Cf. Buchanan and Kuji-Shikatani (2014); Kuji-Shikatani (2015) for experience with this.

in Latin America or Asia (cf. Stockmann, Meyer and Taube 2020b; Stockmann, Meyer and Szentmarjay 2022; Stockmann, Meyer and Zierke [in draft]). The University of Bern, the Universidad Complutense de Madrid and the University of Saarland are among the oldest providers of evaluation masters' programmes in Europe.

The DeGEval is among the few associations which have developed recommendations for academic education and training in evaluation.¹¹ Figure 6.5 a, b and c provide an overview of the competence fields and dimensions.

At Saarland University, in cooperation with the Saarland College of Technology and Economics (HTW), a German-language Master of Evaluation (MEval) course is offered. At the time of writing it was being revised but will be able to be studied again as from the winter semester 2022/23. Another English-language masters course – MABLE – in the blended-learning format, is being offered in cooperation with the Distance and Independent Studies Center (DISC) at the TU Kaiserslautern, and enjoys worldwide popularity (<https://www.disc.uni-kl.de>). Figure 6.6 provides an overview of the principal contents of the MABLE programme.

If we look at these *competence profiles* and *teaching content*, it becomes clear that the work of an evaluator calls for a very wide range of competences, which can only be imparted by education that extends to cover both theory and practice. In order to be able to meet these qualification requirements better than they have been met so far, it is first necessary to systematise the capacity-building measures which are offered increasingly by all kinds of organisations online and on-site, and gear them to general educational goals, in order not just to convey a patchwork of competences that are more or less unconnected. Secondly, in many cases, academic training facilities with modular teaching content and practical relevance are necessary, tailored to the specific conditions of an evaluation.

Praxis-oriented university study courses, which include a period of practical training and lead to an academic title, do not just guarantee extensive education which comprises all the important competences, but also make an essential contribution to the professionalisation of evaluation.

SUMMARY

- ✓ The requirement profile for evaluators is very complex. It comprises not only professional, theoretical and methodological competences, but also organisational, management, social and self-competences and context-related competences.
- ✓ The various different competence catalogues and qualification requirements overlap a good deal when compared internationally, although they do have different areas of emphasis depending on the organisation.
- ✓ To balance out any deficiencies there may be between evaluators in individual areas of competence, it is often a good idea to form an evaluation team.
- ✓ Capacity-building efforts have grown stronger worldwide and the number of academic training amenities has increased.

¹¹ The recommendations are currently in a process of review which has not yet been completed.

Fields of competence and dimensions in keywords

Theory and history of evaluation	
Basic principles	› definitions – evaluation and evaluation research
	› characteristics
	› functions
	› dimensions
Evaluation history	› lines of development
	› national evaluation cultures
	› influences of context factors on the development of evaluation
Evaluation approaches	› theoretical and methodological approaches and models
	› positioning and terms of reference, methodological design
Evaluation standards	› quality assurance
	› communication instrument
	› conflict management, management of evaluations
Methodological competences	
Main features of applied social research and test design	› development and operationalisation of questions
	› planning of empirical investigations, selection and measurement procedures
Data collection, formatting and interpretation	› basic forms of data collection
	› development of data collection instruments
Statistical knowledge	› univariate frequency distributions, cross tables, variance analysis
	› procedures for the measurement of relationships, significance tests
Methodological competences	
Data analysis	› application knowledge in relevant software packages for quantitative and qualitative data analysis
	› coding and recoding
	› data interpretation and reporting
Project organisation	› scheduling, planning and controlling implementation
	› cost planning and cost control
	› introduction to questions of cost-benefit accounting

Figure 6.5a *Fields of competence and dimensions in keywords*

Organisational and subject knowledge	
Organisational knowledge	› organisational concept, organising › action vs. structure
Legal and public administration knowledge	› communication and interaction › introduction to jurisprudence › administrative science › business processes
Specific subject knowledge	› different areas of practice › specific organisational and communication knowledge › specific legal and administrative knowledge
Social and personal competences	
Social competence	› development and arrangement of contacts › adoption of perspectives, empathy › feedback, conflict management
Communicative competence	› communication theory › practice I: listening and speaking › practice II: reading and writing
Cooperative competence	› presenting and moderating › negotiating › cooperation and group working
Self-management competence	› motivation and working style › clarification of mandates, expectations and roles
Learning and problem-solving competence	› reflection and focusing › problem-solving strategies › forms and styles of learning

Figure 6.5b *Fields of competence and dimensions in keywords continued*

- ✓ Thanks to (often still parallel and more or less unconnected) efforts to certify evaluators made by evaluation societies and networks, or in the context of university study courses, the competence profiles of evaluators are being codified and thus demarcated from other professions.

Evaluation practice	
Scope	<ul style="list-style-type: none"> › as a rule 12 weeks › introduction to two fields of evaluation
Quality of trainers	<ul style="list-style-type: none"> › many years of evaluation experience › practical experience in all phases
Evaluation phases	<ul style="list-style-type: none"> › planning (negotiation with clients, evaluation design, project and cost plan) › implementation (data collection and analysis, “breaks” between theory and practice) › presentation of findings (orientation to application, orientation toward political field and “interested lay persons”, compilation and presentation of report) › results check (access to findings from evaluations that have already been completed, application-oriented relevance of evaluation recommendations)

Source: DeGEval 2008, recommendations for education and training in evaluation (<https://www.degeval.org/home/>).

Figure 6.5c *Fields of competence and dimensions in keywords continued*

6.6 BENEFIT AND USE OF EVALUATIONS

6.6.1 Forms of Use

As has become clear from the preceding remarks, the ‘success’ of an evaluation depends on many factors. In the understanding of evaluations represented here, that success can be measured primarily by whether or not it succeeds in generating benefit (cf. Beywl 2001, p. 160; Balthasar 2009, p. 486; Fitzpatrick, Sanders and Worthen 2012, p. 479; Rossi, Lipsey and Henry 2019, p. 310ff.). Even if an evaluation has been conducted very thoroughly in scientific terms and presents valid, reliable, intersubjectively verifiable results, it has failed to achieve its purpose if those results are not then used. From the very beginning, evaluation has occupied itself with the question of what is actually to be understood by the term ‘*benefit*’ of an evaluation.¹² In the 1960s and 1970s, the boom years of evaluation in the USA, it became clearer and clearer ‘that their evaluations did not routinely appear to change’ (Mark 2011, p. 108) or, expressed in even more drastic terms, ‘that evaluations rarely do change policies’ (Højlund 2014, p. 26).

Thus the systematic investigation of the use of evaluations began, as did that of the factors that inhibit it and promote it. Since then – particularly in the USA – countless articles have

¹² For an overview of the topic ‘use of evaluations’ cf. Cousins and Leithwood (1986); Shulha and Cousins (1997); Weiss (1998b); Feinstein (2002); Henry and Mark (2003a, 2003b); Leviton (2003); Mark and Henry (2004); Sandison (2006); Mark (2011); Ledermann (2012); Højlund (2014); Alkin and King (2017); King and Alkin (2019); Patton (2020); Kopischke and Trommeter (2021).

Module 1 Introduction to Evaluation	<ul style="list-style-type: none"> •UNIT 1: INTRODUCTION – WHAT IS EVALUATION? •UNIT 2: THE HISTORICAL DEVELOPMENT AND ROLE OF EVALUATION IN SOCIETY •UNIT 3: EVALUATION APPROACHES •UNIT 4: CLASSIFICATIONS OF EVALUATION APPROACHES •UNIT 5: THE USE AND QUALITY OF EVALUATIONS
Module 2 Conducting and Managing Evaluations	<ul style="list-style-type: none"> •UNIT 1: PHASES OF EVALUATIONS •UNIT 2: THE PLANNING PHASE •UNIT 3: THE IMPLEMENTATION PHASE •UNIT 4: THE UTILISATION PHASE
Module 3 Evaluation Design	<ul style="list-style-type: none"> •UNIT 1: INTRODUCTION TO EVALUATION DESIGN •UNIT 2: REQUIREMENTS, CHALLENGES AND NEEDS •UNIT 3: EVALUATION TYPES •UNIT 4: PRACTICAL EXAMPLES OF EVALUATION DESIGNS
Module 4 Data Collection	<ul style="list-style-type: none"> •UNIT 1: INTRODUCTION TO QUALITATIVE, QUANTITATIVE, AND MIXED METHOD APPROACHES •UNIT 2: INVESTIGATION DESIGN •UNIT 3: COLLECTING INFORMATION •UNIT 4: SAMPLING •UNIT 5: ORGANISATION AND MANAGEMENT
Module 5 Data Analysis Methods	<ul style="list-style-type: none"> •UNIT 1: PREPARING FOR QUANTITATIVE DATA ANALYSIS •UNIT 2: DESCRIPTIVE STATISTICS •UNIT 3: INTERFERENCE STATISTICS •UNIT 4: PREPARING FOR QUALITATIVE DATA ANALYSIS •UNIT 5: QUALITATIVE DATA ANALYSIS METHODS •UNIT 6: CHALLENGES IN USING QUALITATIVE DATA
Module 6 M&E in Organisations and Economic Evaluation	<ul style="list-style-type: none"> •UNIT 1: BASICS OF ECONOMIC EVALUATION •UNIT 2: FOUR TYPES OF ECONOMIC EVALUATION •UNIT 3: INTRODUCTION TO ORGANISATIONS AND NETWORKS •UNIT 4: M&E IN ORGANISATIONS •UNIT 5: M&E OF ORGANISATIONS AND NETWORKS •UNIT 6: ORGANISATIONS AND NETWORKS AS PROFESSIONAL AGENCIES FOR M&E
Module 7 Key Communication Qualifications in Evaluation	<ul style="list-style-type: none"> •UNIT 1: THEORETICAL BASICS OF COMMUNICATION •UNIT 2: READING; WRITING AND PRESENTING •UNIT 3: FACILITATION •UNIT 4: REPORTING AS A COMMUNICATION STRATEGY •UNIT 5: NEGOTIATION •UNIT 6: MEDIATION AND CONFLICT RESOLUTION IN EVALUATION PROCESSES
Module 8 Elective Module	<ul style="list-style-type: none"> •EVALUATION OF THE SUSTAINABLE DEVELOPMENT GOALS

Source: TU Kaiserslautern, Saarland University: Module Handbook Master (Blended Learning) of Evaluation (MABLE). Handbook 2.1 (May 2021). Available at: <https://www.zfuw.uni-kl.de/sites/default/files/media/user-68/files/Module%20Handbook%20%28Final%20May%202021%29.pdf>.

Figure 6.6 Contents of Masters (Blended Learning) of Evaluation (MABLE)

appeared on the subject, so that Højlund (2014) comes to the conclusion: ‘Evaluation use is probably the most researched theme in the literature on evaluation’ (cf. also Henry and Mark 2003a, p. 294). That is not surprising, because on the one hand evaluation kicks off with the aspiration ‘that evaluation should have consequences’ (Mark 2011, p. 108) and on the other, those who finance evaluations expect to get useful results for their money.

Research into the use of evaluations is confronted with several challenges:

- (1) As already shown in the preceding chapters, there is no uniform understanding of evaluation, so that many very different things take place under this label.
- (2) It follows from that that the benefit obtained from such a great variety of activities can also come in many different guises.
- (3) It thus comes as no surprise that there is no clear definition of benefit, let alone a uniform one, and that various different dimensions are covered by this term.
- (4) The result of an assessment of whether and to what extent a benefit – however it is defined – has come about as the consequence of an evaluation, depends – and who should know that better than evaluators? – on the criteria applied. For this too, there is a wide range of suggestions.
- (5) One astonishing thing is that many evaluators of programmes call for a theory of change and scrutinise it critically in evaluations, but do not develop a theory of change themselves for the achievement of their uppermost aim, i.e., that of generating useful findings with an evaluation.

In spite of these difficulties, the past five decades have seen tremendous growth in the state of knowledge about the empirical benefit of evaluations, about the taxonomies, about what is to be understood by them and about the influencing factors, and all of this has found its way into the evaluation standards.

In the early years, use was above all defined as a direct impact on decisions (cf. Herbert 2014, p. 389). Very soon, however, it became clear that this concept falls short of achieving its aim.¹³ As far back as the beginning of the 1980s, distinctions were made between *three forms of use*, and these are still applied today (cf. Stockmann 2008, p. 276; Mark 2011, p. 108f.; Saunders 2012, p. 425; Alkin and King 2016, p. 574ff. and others):

- (1) *Direct (instrumental) use*: This term denotes the direct, practical use of evaluation results by the clients and other stakeholders. For example, one such kind of use occurs when results are put to use in decision-making, programmes rearranged in accordance with the evaluation recommendations or strategies, and communication relationships etc. changed.
- (2) *Conceptual use*: Conceptual use occurs when evaluation results change the general way those involved think about certain problems, for example if the evaluation enables fundamental insights to be gained about the way a programme functions and the impact it has.

Sometimes a distinction is made here to *convictional use*. This sets in when evaluation results are used to reinforce ‘political’ positions or cast doubt on them. This is the case, for example, when the findings of evaluations are able to refute firmly anchored positions which were no longer being challenged.

- (3) *Political or symbolic use*: This denotes a use of evaluation results that only serves to confirm existing positions before the evaluation. In other words, use is only made of the evaluation results if they correspond to one’s own fixed preferences, predispositions or plans, and only then. Findings used in this way include those of evaluations which

¹³ On the historical development of research into the use of evaluations cf. Herbert (2014); Alkin and King (2016), and others.

are only intended to serve to legitimise decisions that have already been made after the event.

Yet finer definitional distinctions are also made: for example, between tactical, political, symbolic, legitimative use etc.

Alkin and King (2016, p. 575) point out that as long ago as the 1970s and 1980s a number of authors¹⁴ observed ‘that the evaluation process itself could have impact’. However, it was Michael Patton (1994) ‘who named the concept and brought the idea of process use to the attention of the evaluation community’ (ibid.). Patton (1997, p. 90) writes:

Process use refers to and is indicated by individual changes in thinking and behaviour, and program or organisational changes in procedures and culture that occur among those involved in evaluation as a result of the learning that occurs during the evaluation process. Evidence of process use is represented by the following kind of statement after an evaluation: the impact on our program came not just from the findings but from going through the thinking process that the evaluation required.

Learning processes that alter the way people think and act because they take part in an evaluation can occur in all three of the above-mentioned forms of use. Alkin and King (cf. 2016, p. 575) thus argue that *process use* should not be considered as an extra category, but as a transverse dimension. Whether or not this is true, it is important to note that evaluations can be useful both on account of their findings and on account of their planning and implementation processes.

As a consequence of this insight, evaluations are now being organised increasingly as learning processes whose primary aim is to learn in the planning and conducting of an evaluation and not from the evaluation results it produces. This change in strategy has led to the development of an increasing number of participative approaches (cf. Chapter 2.2).

However, dissatisfaction with the *inexactitude of the term ‘use’* has persisted. There has furthermore been criticism to the effect that whilst various types of use have been defined, there has been no explanation of how they can be achieved. In addition, the necessity for a follow-up concept has been substantiated by saying: ‘that the term use tends to underestimate the impact of evaluation by emphasising results-based use and that use suggests an intentionality, immediacy, and directness that may not always exist’ (Herbert 2014, p. 393).

The newly created concept of ‘*evaluation influence*’ was seen as an extension of the ‘*use*’ concept, in order to be able to record the impacts of evaluations more comprehensively (Kirkhart 1995, 2000; Henry 2000; Henry and Mark 2003b). However, the expansion of the term to include evaluation effects which ‘plausibly lead toward or away from social betterment’ (Henry and Mark 2003a, p. 295) indirectly, unintentionally and in the long term, did not lead to a greater degree of precision in the definition of the term, so that it was not long before criticism came (cf., for example, Alkin and Taut 2003; Hofstetter and Alkin 2003; Cousins 2004; Patton 2008). In a meta-analysis of 28 studies of evaluation influence, in any case, Herbert (2014, p. 389) arrives at the devastating conclusion that ‘Many of the studies reviewed offered vague and inconsistent definitions and have applied influence in an unspecified way in research.’

¹⁴ Rippey (1973); King and Pechman (1984); Greene (1988); King (1988); Cousins and Earl (1992).

Category of Misuse	Form of Misuse
When the evaluation process is good	
Commissioning	User commissions evaluation for political show or as a delaying action (symbolic use) User commissions evaluation to justify decisions already made and not open to possible changes (legitimative use)
Process	User subverts the evaluation process (e.g., by cutting funding for it, by limiting access to data sources or existing data)
Findings	User modifies data or report content or cherry picks content intentionally User actively misrepresents evaluation information (e.g., actively changing it, distributing incomplete results) User purposefully ignores evaluation findings for personal gain User actively uses evaluation information known to be inaccurate or invalid

Source: Alkin and King (2017, p. 441).

Figure 6.7 *Types of evaluation misuse*

Recently, Alkin and King (2016, 2017) and King and Alkin (2019) have written up the current state of knowledge about the theories on ‘evaluation use and influence’ in three essays that build on one another, and endeavoured on this basis to create a universal operational definition of ‘evaluation use’, though it must also be said that it is not completely convincing (Patton 2020).

In the conceptualisation of the use of evaluations, the negative side of the coin – their misuse – ought not to be forgotten. Some of the cases listed in Figure 6.7 by Alkin and King (2017, p. 14) have already been discussed above, so that further explanations can be dispensed with here.

An interesting change of perspective is proposed by Højlund (2014), who points out that as a rule evaluations take place in an organisational setting. Typically, they are commissioned by organisations and conducted by organisations. For that reason, he recommends that the evaluators pay more attention to the *organisational context* of evaluations to increase their benefit: ‘ (...) evaluation use or non-use are likely to be explained better by understanding the organisational context in which the evaluation takes place, rather than the evaluation’s attributes or other “micro-level” explanations of conditions and factors’ (Højlund 2014, p. 38).

This highlighting of the importance of the organisational context of evaluations complements the organisational model of the CEval approach presented in Chapter 2.2.2. In that model, the argument is put forward that interventions (projects, programmes etc.) are carried out by organisations, and that because of that the latter’s performance capability also plays a decisive role in the success of the programme. Thus (project / programme) interventions can bring about impacts both inside and outside the organisation in the intended areas of impact (sectors, subsystems). Organisations are thus not only a decisive factor in the success of the programme, but also – as Højlund emphasises – important for the use of evaluation results (cf. also Fitzpatrick, Sanders and Worthen 2012, p. 481).

SUMMARY

- ✓ Evaluations can generate benefit in various different ways. Evaluation findings can be used directly (instrumentally), conceptionally or for symbolic policy-making.
- ✓ Furthermore, participation in an evaluation can be useful for the actors involved. That is the case if the evaluation contributes (i.e., instrumentally) to making changes in the evaluand (e.g., programme), or changes in the way the actors think about the programme or evaluation (conceptionally) in itself. This type of use is referred to as process use.
- ✓ Evaluations can not only generate positive benefit; they can also be misused in various ways.

6.6.2 Factors that Increase Benefit

The many studies on the benefit of evaluations have made an essential contribution to extending the concept of use to include new perspectives and mapping out the factors that influence that use theoretically and empirically.

The following factors are listed as *increasing the benefit* of evaluations (cf. Cousins and Leithwood 1986, p. 331ff.; Shulha and Cousins 1997; Fitzpatrick 2012, p. 485ff.; Strobl, Lobermeier and Heitmeyer 2012, pp. 12, 61; Alkin and King 2017, p. 445f.; Rossi, Lipsey and Henry 2019, p. 311ff.; Owen 2021, p. 113ff.). A number of success factors are grouped around the competent planning and conducting of an evaluation:

- analysis of the evaluation object in its context (policy and decision-making structures, stakeholders' interests, organisational context etc.);
- development of a theory of change that is appropriate to the problem analysis / deficiency analysis;
- use of an evaluation approach that is appropriate to the evaluation purpose and design, and adequate methods of data collection and data analysis;
- scientific quality of the results (validity, reliability, objectivity);
- usefulness of the findings for the information and decision-making requirements of the clients and principal stakeholders;
- clarity of the report and transparency of the recommendations;
- timely delivery of the evaluation findings and submission of the evaluation report;
- integration of the stakeholders in the evaluation process;
- frequent and regular communication that caters to the information requirements of the stakeholders.

Other success factors can be ascribed to the *clients* and their *organisational context*:

- clear and precise formulation of the information requirements of the clients of an evaluation;
- provision of sufficient resources for the professional implementation of the evaluation;
- interest of the members of the organisation at the various levels (management, programme management etc.) in the findings of the evaluation;
- acknowledgement of the importance of the evaluation results by the members of the organisation in competition with other sources of information;

- existence of an organisational structure that is open to criticism and oriented toward learning;
- general readiness of individuals to come to terms with criticism, learn from mistakes and make changes.

A few other features are associated directly with the *evaluator*:

- competence of the evaluator / evaluation team to plan and conduct evaluations correctly and professionally;
- reputation of the evaluators (based on their previous achievements);
- credibility (based on criteria such as objectivity, incorruptibility etc.).

Not usual so far, but worth giving some thought to, would be the development of a *theory of change for the evaluation* (alongside the programme theory) together with the client (stakeholders) in the planning phase, to make it clear where an evaluation should begin to enable changes to come about.

SUMMARY

- ✓ In the last 50 years, research into the benefit of evaluations has mapped out many factors that have an influence on their use.
- ✓ Fundamentally, a distinction can be made between success factors, which (1) relate to the planning and implementation of evaluations, (2) are associated with the clients and their organisational context and (3) are connected with the evaluators themselves.
- ✓ It is not usual to develop a theory of change for evaluations (complementary to a programme theory) that describes how changes should be brought about by an evaluation.

6.6.3 Usefulness Standards

Many of these insights have found their way into the *standards* of the evaluation societies or were inspired by them. King and Alkin (2019, p. 432), for example, draw attention to the fact that ‘evaluation use’ has a role to play both in the Program Evaluation Standards – first published in 1981 and revised in 1994 and 2011 – and in the AEA’s Guiding Principles for Evaluators (2018). The same applies to the DeGEval’s ‘Standards for Evaluation’. The DeGEval has defined *eight usefulness standards*, which are intended to ensure ‘that the evaluation is geared to agreed and clarified evaluation purposes and, as far as possible, to the information requirements of the intended users’ (DeGEval 2016b, p. 34).

USEFULNESS

N1 – Identification of stakeholders

Those involved in and those affected by the evaluand or evaluation should be identified ahead of the event so that their interests and information requirements can be made clear and taken into account as much as possible in the evaluation design.

N2 – Clarification of evaluation purposes

It should be determined clearly what purposes are being pursued with the evaluation, so that those involved and those affected can take a stance and the evaluators can pursue a clearly defined work assignment.

N3 – Competence and credibility of the evaluator

Those conducting evaluations should be professionally and methodologically competent, so that the highest degree of credibility and acceptance can be achieved for the evaluation and its findings.

N4 – Selection and scope of information

The selection and scope of the information gathered should make it possible for the questions on the evaluation object that require investigation to be answered adequately. They should also cater to the information requirements of the clients and other parties who are involved and/or affected.

N5 – Transparency of values

Values of those involved and those affected such as manifest themselves in their points of view and the assumptions they make, and have an influence on the evaluation and the interpretation of its findings, should be documented in a transparent way so that the findings can be classified better.

N6 – Completeness and clarity of reporting

Evaluation reports should provide all the essential information and be comprehensible and transparent for their audiences.

N7 – Timeliness of evaluation

Evaluation projects should be commenced and completed in such a way that their findings can be integrated in pending decisions, improvement or other processes of use.

N8 – Use and usefulness of evaluation

The planning, conducting and reporting of an evaluation should encourage those involved and those affected to contribute to the evaluation and make use of its findings.