Preprint

Evaluation of the Educational/Training activities/programme and impact measurement

Marcus Flachmeyer

Abstract

This article provides a basic introduction to the essential aspects of the evaluation of educational and training activities and programmes in youth work. First, the ethical challenges of an evaluation and the importance of clarifying the objectives are discussed. The article highlights the advantages and disadvantages of external and internal evaluations, as well as the two forms of evaluation: 'summative' and/or 'formative' evaluation. Another focus is on the practical planning of an evaluation in seven steps. Finally, four evaluation models are presented that can be applied in non-formal education (Kirkpatrick, Kaufman, Phillips, CIRO), and three models of programme evaluation are mentioned that fit well with the self-image and mission of youth organisations (Stake, Cousins & Earl, Fetterman). The article emphasises the need to understand evaluation as an integral part of youth work in order to continuously improve pedagogical practice.

Flachmeyer, M. (2022). Evaluation of the Educational/Training activities/programme and impact measurement. In *STREAM Entrepreneurship approach implementation in youth work. Guidelines for youth workers* (S. 40–59): Erasmus+ project partnership STREAMpreneur.

Evaluation of the Educational/Training activities/programme and impact measurement

Marcus Flachmeyer

Introduction

Today, not only the needs assessment, the planning/organisation and the implementation of educational activities are part of the methodological spectrum of youth work, but also their evaluation and improvement. In practice, however, the focus is on "doing", understood as planning, organising and implementing. Evaluation plays a rather subordinate role; it appears as satisfaction surveys and is a well-proven way of giving all young people a chance to have their say. On the other hand, the evaluation of educational programmes, usually initiated by the funder or/and the executing agency, does not exactly trigger enthusiasm among many professional and voluntary youth workers, but also among the young people themselves. "Well, if that's what it takes..." is often the reaction.

This paper does not aim to provide the definitive recipe for generating enthusiasm for evaluation in youth work. This is in the hands of the professional and voluntary staff who have to decide again and again whether and in which way the effects of their pedagogical work, educational activities and educational programmes are systematically recorded, presented and communicated. This article provides some guidance for such decisions. It first discusses the ethical challenges of an evaluation and the clarification of the target functions, then goes into the question of external or internal evaluation, before discussing the two forms of evaluation "summative" and/or "formative". The practical planning of an evaluation is then in focus, supplemented with four evaluation models applicable in non-formal education and three programme evaluation models that are very much in line with the self-image and mission of youth work organisations. In these explanations - as far as space allows - the implementation possibilities are not only addressed in organisations working with young people, but also specifically in STEM education.

Definitions

First, the terms already used in the chapter title are determined in order to create a common starting point for reading this chapter. These determinations are of a rather formal nature, they will be concretised in the further course of the chapter.

Evaluation is to be understood both as a process and as a result, Scriven calls evaluation a "process of determining the merit or the worth or the significance of something; or the product of that process". (1981, p. 53).He defines the central terms contained therein: "Merit" is the "intrinsic value as opposed to extrinsic or system-based value/worth" (p. 94), "Worth" is "the system value by contrast with intrinsic value" (p. 167) and "Significance" is the "overall, synthesised conclusion of an evaluation, may relate to social or professional or intellectual significance." (p. 145). The objects of evaluations can be diverse, including programmes, projects, services such as training, products, personnel, organisations, data, theories, policies.

Educational activities and **educational programmes** span the entire spectrum of informal and non-formal education. In this context, a programme is to be understood as the larger and more complex format, namely a whole bundle of interrelated measures or activities with a specific long-term goal, usually conceived, developed and implemented top-down. The Joint Committee on Standards for Educational Evaluation (JCSEE) defines a programme as an "orchestrated initiative that dedicates resources and inputs to a series of activities intended to achieve specific process, product, services, output, and outcome goals" (Yarbrough, et al., 2011, p. 291). Educational activities, on the other hand, are the smaller and less complex format; they may or may not be part

of educational programmes. Educational activities e.g. in a youth centre often have a rather singular character, they react in the short term to the visible or expressed needs of young people, their strategic dimension often remains unclear. However, they can also have a clear strategic dimension, e.g. when it comes to the competences of leaders in youth associations.

Training activities and **training programmes** also differ from each other in the characteristics just described: training programmes are a bundle of interrelated training measures or activities with a specific long-term goal (see above). Training is to be understood as a specific form of non-formal education in youth work. Training, according to a definition from the European youth work discussion context, aims to empower young people through the development of knowledge and competences for personal and (increasingly) professional life, where the training content is intrinsically relevant and useful (Council of Europe, 2021). On the provider's side, the teaching objectives (and possible learning goals) for such trainings are usually formulated, the training content is methodologically-didactically prepared, the training is structured in terms of time and a trainer is provided.

Impact measurement is a variant of evaluation that has always had a special place. "What's interesting is what comes out at the end ..." - and particularly interesting are the effects. Scriven defines impact evaluation as follows: "An evaluation focussed on outcomes or pay-off rather than process delivery or implementation evaluation" (1981, p. 74). Outcomes are usually the post-treatment effects, there are often effects dunng treatment.

Evaluation and responsibility/ethics

Regardless of whether it is the evaluation of a single training or a complex educational programme, those who initiate or conduct an evaluation should be aware that with an evaluation they are entering a field of expertise that is characterised by extensive theoretical foundations and diverse practical experiences. As can already be seen in the basic definition, an evaluation always involves value judgements (which in turn can of course be assessed again). Evaluations should always be set up with the awareness that value judgements require an even higher degree of responsibility for the procedure as well as for the results compared to mere descriptions or analyses.

In any case, reservations about evaluation are very much on the rise, as they touch on fundamental questions of power and ethics. Such questions are not unfamiliar to the evaluator community, on the contrary. Evaluation standards are, among other things, an attempt to create more security for all those involved in this demanding terrain. The Joint Committee on Standards for Educational Evaluation (JCSEE), which codified the basic consensus of the evaluation discipline with evaluation standards for the first time in 1981, is leading the way here. There are thirty standards in the current version of the JCSEE Handbook of Evaluation Standards (Yarbrough, et al., 2011), divided into five dimensions: Utility, Feasibility, Propriety, Accuracy and Accountability. These standards aim at responsible evaluation of educational programmes, projects and materials. They are to be understood as "guardrails" that can point the way towards a responsible evaluation project when formulating a concrete evaluation project.

These five dimensions or categories of good evaluation quality, characterised by care and fairness, contain thirty standards. Each of the thirty standards is described in the JCSEE manual in such a way as to first set out its meaning and purpose, followed by very detailed guidelines for applying the standard, a list of frequent mistakes and one or more illustrative examples of the practical application of the standard. This structure underlines the intention to actually influence the practice of evaluation and lead to responsible and ethical evaluation projects.

When an evaluation is planned and carried out in the field of child and youth work, the questions of power and ethics that are inherent in any evaluation arise once again in a very special way. Evaluation is called upon to address young people in a way that corresponds to their cognitive and emotional stage of development and

takes into account their specific need for protection. Evaluation, understood as practice research or applied research, can benefit from a discourse in the research community that has been going on for several years, e.g. from the ERIC project. The project "Ethical Research Involving Children" (ERIC) sees itself as a forum and initiator for the research community with the aim of ensuring that the rights of children and young people are also fully respected in research processes. The ERIC Guidelines call on researchers to address the complexity of ethical issues and their values, attitudes, beliefs and assumptions, and to recognise the ways in which these influence the decisions they make in the research process. (UNICEF: Office of Research - Innocenti ERIC, 2013).

Target functions of evaluation

It has proven to be very useful to start conceptual considerations with the question of what the evaluation should serve. A simple model of goal functions outlined by Stockmann (2004) can be helpful here. Stockmann fundamentally distinguishes between four goals: (1) gaining knowledge; (2) exercising control; (3) creating transparency to enable dialogue; and (4) documenting success (legitimation).

Goal function 1: Gaining knowledge (knowledge function)
Evaluations should provide insights that are of interest, in some way valuable and useful to the clients of the evaluation and the target groups of the programme (the object of evaluation).

Goal function 2: the execution of control (control function)
Stockmann assumes that while the primary interest in evaluation is to gain knowledge, he also sees the interest, especially of clients, in executing control. This is very familiar to organisations that work with public funds, since interim and final reports or monitoring discussions are always about the question of whether and how the submitted plan is actually realised, what difficulties there are and how they are dealt with. "I.e. every evaluation is directly or indirectly connected with a form of control." (Stockmann, 2004, p. 4).

Goal function 3: the creation of transparency to enable dialogue (learning function).
Stockmann (2004) always sees evaluations as an opportunity for transparency and dialogue, a concern that is particularly essential for work with children and young people. But it is generally true that if the findings of an evaluation are made transparent, they enable a dialogue between different stakeholders such as the donors, the implementing organisation, the target groups and other participants and affected parties. This always provides a basis for collective learning.

Goal function 4: the documentation of success (legitimisation function)
Evaluations very often aim to legitimise a certain use of resources (input) and a certain procedure (process) and to relate them to the result (output/oucome/impact). Obviously, this is not always trivial and complex evaluation questions also require very elaborate evaluation designs (which in turn cost money). The results can be used by implementing organisations (as well as donors) to legitimise their own work.

Stockmann (2004) also points out that evaluations are very often ascribed "tactical" functions (p. 4). One speaks of tactical functions when the results of evaluations are only to be used to legitimise certain decisions that have already been made, or because it has to be easy or is chic. In this context, he refers to Pollitt, who already lamented a "pathological side" in 1998: "Unfortunately, however, there is a pathological side to these developments. There are temptations for politicians to use evaluations as baubles or as bolsters-as decorative symbols of modernity or simply as reinforcements for courses of action they had already decided upon for other reasons" (p. 223).

Internal and external evaluation

Especially in the evaluation of more complex educational activities such as educational programmes or even somewhat larger, project-based educational activities, the question often arises very quickly in practice as to whether an evaluation contract is awarded externally (external evaluation) or whether the evaluation

can/should be planned and implemented with the institution's own staff (internal evaluation). The answer to this question is often determined by trivial factors. The reason for the decision for an internal evaluation is often simply the lack of money for an external evaluation. If there is money, the commissioners of an external evaluation usually expect more distance, a more comprehensive view and more objectivity; at the same time, they sometimes want to prevent the impression that they are not really interested in an objective evaluation from the outside.

Internal evaluation

With internal self-evaluation, the design, planning and implementation (and meta-evaluation if necessary) is entirely in the hands of the own organisation. This has a lot of advantages. For example, it can be assumed that access to the field of practice to be evaluated is possible without problems and that the evaluation can be fitted methodologically and temporally tightly into the pedagogical processes, which facilitates data collection. Conceptually, the organisation itself has a deep understanding of the object to be evaluated. On this background, it is able to ask the right questions, to classify the collected data and to interpret them close to its own practice.

A possible disadvantage could be the lack of distance to the subject matter. This possible disadvantage has at least two facets, (1) a research paradigmatic and (2) a professional facet. In terms of research paradigm, it is often argued that a lack of distance from the subject matter impairs the objectivity of the evaluation or even makes it impossible. Here it must be considered whether an objectivist understanding of evaluation/research is actually appropriate to the object and the project, or whether other approaches, e.g. reconstructive social research or action research, do not fit better. The second facet, that of professionalism (of the organisation), also needs to be clarified. An essential characteristic of professionalism is reflexivity, i.e. the ability to consider and classify one's own actions and to think in terms of alternatives; it is in fact the very first prerequisite for internal evaluation. It is likely to be more and more given in the course of professionalisation of work with young people.

Another characteristic of internal evaluation, which is possibly seen as a disadvantage, is the related qualification requirement for the organisation. Internal evaluation, especially when it comes to the evaluation of educational programmes or projects, presupposes specific, disciplinary knowledge about evaluation. However, it should be noted that some organisations now have such a capacity, in specific organisational units or at management level. The capacity for self-evaluation (as a variant of internal evaluation) at the level of pedagogical staff has been systematically addressed in recent years in the course of the discussion on the effectiveness of pedagogical action in youth work organisations (e.g. Abele & Erdmann-Kutnevic, 2016; Council of Europe, 2018; Council of Europe Youth Work, 2021; König, 2007; SALTO Training and Cooperation Resource Centre, 2021; YouthLink Scotland, 2021).

External evaluation

In an external evaluation of mostly larger, complex educational programmes, the organisation places itself in the hands of external specialists and rightly expects highly specialised declarative and procedural knowledge. This is an expected advantage of external evaluation, especially since it can be assumed that the external institute and its evaluators are independent and implement the standards of responsible evaluation. If the donor expects an external evaluation or if the commissioning organisation lacks its own capacity, external evaluation is the procedure of choice.

However, choosing an external evaluation does not mean that the client can sit back and just pay. A successful external evaluation depends on the cooperation of the client. A good evaluation institute develops the evaluation design together with the client, plans the procedure together with the client and implements it in close coordination and with the necessary support of the commissioning organisation. It prepares - according to the agreement - interim reports and the final report and, if necessary, presents the results. The own organisation is thus relieved and at the same time methodically guided through the evaluation project.

Nevertheless, possible risks and problems of external evaluation must be taken into account. It is not uncommon that an external evaluation is not necessarily welcome at the working team level and triggers defensive reactions. This may not only affect the quality of the data, but the evaluation may also not fit into the pedagogical processes in the way that might be desirable. In addition, there is the risk of misinterpreting the data due to lack of insider knowledge (depending on the evaluation design) and drawing conclusions that are not only wrong but also unusable. These potential problems have long been known in the evaluation discussion and were, among other things, the impetus for the "fourth generation evaluation" (Guba & Lincoln, 1989), which includes the stakeholders and especially those who are "researched" in the interpretation, classification and evaluation. Behind this concept is, on the one hand, a constructivist understanding of knowledge and science and, on the other hand, the insight that an evaluation is only useful if the stakeholders participate in the evaluation in a co-creative way, make its results their own and then, intrinsically motivated, take them into account in practice as part of their professional self-understanding.

Formative and summative evaluation

The terms formative and summative evaluation were coined by Scriven (1967), dating back to an exchange with Lee Cronbach, who took a completely different view of the goal function of evaluation than Scriven himself. Cronbach (1963) held that the evaluation of an educational programme is particularly useful if it gives indications for improvement already during the implementation of the educational programme (and not only at the end). In his view, such benefit is far more valuable to teachers than the after-the-fact evaluation of the educational programme. Scriven (1967) took up this argument in the context of a discussion on the roles and goals of curriculum evaluation. Thus, he also considered it useful to test already during the development work and to process the findings in the development process. He called this form of evaluation formative evaluation (p. 41). An evaluation that retrospectively assesses a curriculum (or any other product) as a whole, he called summative evaluation.

Evaluation models and evaluation practice today often make use of both forms of evaluation. In the course of the educational programme, the educational project or the educational activity, data are collected at several sensibly selected points in time with the aim of positively influencing the implementation process. The data is processed and, in a participatory approach, also comprehensively reintroduced into the implementation process of the measure. Such a participatory approach is a must, especially when working with children and young people (cf. Hart, 2008, among others), but it is not a foregone conclusion. The results must always be prepared in a way that is appropriate for the target group and meaningfully linked to the main pedagogical process, an approach that is not always easy to implement in practice and must be carefully planned. Once the educational programme, the educational project or the educational activity has been completed, data is collected again, directly after the end and/or with a time lag. They are the basis for the summative evaluation, which assesses the measure as a whole in retrospect; findings from the formative evaluation are usually included.

Planning an evaluation

The number of manuals and checklists that can help in planning and implementing an evaluation project is now almost unmanageable. On the one hand, this reflects the extent to which evaluation has moved beyond the field of education programmes, which has been exclusive for many years, into other areas such as health, environment, transport, society, and so on. On the other hand, it shows that evaluation has developed and differentiated from within in the way it is planned and implemented, but that it is also subject to external trends, such as those visible in project management. Whereas in the 1980s and 1990s careful and lengthy analyses, definitions and planning followed by a very linear implementation of the project plan were the "state of art", today flexibility, agility and collaboration are the order of the day (Michaelides & Antonacopoulou, 2021). All in all, there is not only an unmanageable number, but also a variety of approaches and not one silver bullet.

Against this background, which path an organisation in youth work and its pedagogical staff take is also and especially a question of organisational culture, the organisation's own abilities and practices. Evaluation must "fit" and embed itself so that it can fulfil its essential function of improving pedagogical practice. In this respect, the following planning steps are to be understood as recommendations, which always need to be adapted to one's own needs, but without compromising the evaluation quality.

Step 1: Clarify reasons, goals and subject of the evaluation, establish consensus

It is helpful if the different stakeholders and users are clear about what the evaluation is actually about right at the beginning. This urgency exists not only in the evaluation of educational programmes and projects, but - broken down accordingly - also in the evaluation of a training course. Even with a comparatively small unit, legitimacy and transparency are needed. This already starts with the trigger. "Who came up with this idea?" - "It's a condition of the funding programme" can then be the clear answer in the case of an educational project, for example. Most of the time, the purpose of the evaluation (see above) is already clear from the trigger. "What is actually the thrust?" If it is about control, it is helpful if this is also made clear in no uncertain terms. But perhaps difficulties have arisen and the cause of these difficulties is not entirely clear to the pedagogical staff member in charge, in which case the insight function is in the foreground. And the object of the evaluation should also be narrowed down directly. "What is to be looked at?" This focus helps to be well prepared for the next step.

Step 2: Clarify capabilities and resources

Once a consensus has been reached on what the evaluation should be about, it can be quickly clarified whether the organisation has the necessary capabilities and resources for an evaluation project. It hardly matters whether the evaluation is of an educational programme, an educational project or an individual educational activity such as a training course. Much more decisive for the assessment of one's own capabilities and resources is an idea of what the evaluation is supposed to be about (and what methodological and instrumental effort this will probably require).

If the organisation has the impression that it has the professional capabilities to plan and conduct an evaluation itself, it will also be in a position to quantify the internal resources required. An initial budget should be outlined at this point, on the basis of which it can be assessed whether an internal evaluation is in a positive relationship to the expected results. If the evaluation is of an externally financed educational project or activity, it may be possible to finance the calculated expenditure there.

If the organisation has the impression that it does not have the necessary professional capabilities itself or prefers an external evaluation for other reasons, it should have an offer made by a professionally experienced institute. A well-made offer will not only give an initial indication of the institute's approach, but also information on the fee and roughly the type and extent of the required participation.

Once this step is taken, the organisation has sufficient clarity regarding its own capabilities and resources. In this context, the question of an internal or external evaluation has also been clarified. Now the available information should be sufficient for a decision for or against an evaluation.

Step 3: Build working capacity

Once this decision has been made, there is a lot of work to be done, for which human and possibly material resources need to be made available. Depending on the scope of the evaluation project, it may only need a few days of work by a single person, but perhaps a project team will now start work, which must be related to the organisation. If it is an external evaluation, the relationship between the organisation as the client and the external institute as the contractor must be contractually fixed. But also in the case of an internal evaluation, the individual person or the project team needs a clear legitimisation and clear rules on the part of the

organisation. The internal principal must be clearly defined, and the relationship to a conceivable steering group or advisory board must be clearly regulated.

Step 4: Set up the evaluation concept

Although it is still common in the practice of calls and tenders that the development of an evaluation concept, the development of instruments and the planning of implementation take place before the approval or commitment, this is neither resource-saving nor efficient. The decision to plan and implement an evaluation should be made now in order to actually have the resources to develop an evaluation concept in close consultation with stakeholders and participants.

But what is actually meant by an evaluation concept? In the German-language glossary of evaluation, an evaluation plan (alternatively: evaluation concept) is defined as:

> A detailed, consistent and written outline of the intended procedure of a specific evaluation. It presents the chosen evaluation approach, the steps to be taken and the evaluation methods to be used (e.g. data collection plan). It specifies responsibilities, reporting procedures and the intended use of the evaluation. The planning is based on the agreed purpose of the evaluation and the evaluation questions. (Eval-Wiki: Glossary of Evaluation, 05.11.2020).

With regard to the special field of use, the work with young people, the evaluation concept should additionally specify in which way the cognitive and emotional developmental level of young people and their specific need for protection is taken into account.

Once this step has been taken, a framework is available for all participants and stakeholders that provides information on the methodological decisions and the procedure.

Schritt 5: Develop/assemble instruments

Once the framework is set, the methods mentioned in it can be instrumented. The first step is data collection, followed by data processing/presentation and data evaluation. In principle, the entire range of empirical social research is available for this purpose, including the broad field of quantitative, qualitative and mixed-method research.

The selection or development of survey instruments requires that the evaluation questions are broken down to the level of research questions. Research questions are much more specific and can focus the research process. Especially in the case of a quantitative impact measurement, they ideally include already introduced constructs such as self-efficacy expectations, willingness to take risks, need for cognition, etc. and are thus the access to already validated item scales. Alternatively, in a quantitative research approach, indicators and items still need to be developed.

Once the research questions are clear, questionnaires, observation forms, interview guides, discussion guides, etc. can be selected or developed. If existing instruments are used, legal rights must be clarified beforehand. The instruments must meet specific validity criteria. The instruments as well as the medium must take into account the respondents with their cognitive, motivational and cultural characteristics.

The processing of data, especially qualitative data, requires a considerable investment of time and an appropriate choice of technical tools. Quantitative data can be processed to an astonishing extent even with MS-Excel; a qualitative content analysis of the results of a group discussion, a narrative interview, a sociodrama, etc. can be done within the framework of a small to medium evaluation project even without analysis software such as MAXQDA. Once the data have been processed, it is important to remember to present the collected data in such a way that they arouse interest, are comprehensible and are experienced as useful. This target group-specific presentation costs time (and thus money).

The analysis and evaluation of the data should be planned in such a way that the stakeholders and participants are encouraged by the evaluators to contribute their interpretations and classifications. This can happen in workshops, one-on-one interviews, in the results report via the pdf comment function, and so on. However, it should not be underestimated that this participatory evaluation requires stimulating formats and resources.

Step 6: Implement the evaluation

The implementation of the evaluation concept and its instruments and procedures needs one thing above all: trust, transparency and commitment. Important promoters are the staff members who give full importance to this evaluation project, communicate it credibly and thus act as role models. It is also beneficial when young people who are key players make a strong case for it, e.g. in social media. Depending on the size of the evaluation project and its importance for the organisation, great emphasis should be placed on communication via different media.

Step 7: Fix and use the results

The evaluation concept has already specified the reporting procedures and the intended use of the evaluation, including the final reporting and the use of the final results. In a good evaluation concept, these measures are designed in such a way that they effectively and efficiently support the goal function(s) of the evaluation project. However, a certain degree of uncertainty remains as to whether these measures can be implemented as successfully as intended.

A structural challenge is the time gap between the end of the educational activity and the processing of the results of the summative evaluation. The processing of the results in a report, podcast or videocast etc. requires a certain amount of time in practice, participatory formats such as a results workshop for stakeholders and participants are sometimes difficult to realise because the young people are no longer approachable and are already focused on other things.

If the stakeholders and participants are actually to take note of the results, their commitment to the organisation and to the educational activity/project is a prerequisite. Evaluation approaches with a maximum degree of participation, such as empowerment evaluation (Fetterman, 1994), are also effective. They favour not only the knowledge of the results, but also the utilisation in the sense of an improvement of the product..

Application of existing evaluation models to non-formal education

When designing one's own evaluation project, no matter how small, it is very helpful to fall back on proven evaluation models. When we talk about models here, we are not talking about mathematical models that are used to establish theories, but about concepts and structures with which the authors approach the object to be evaluated and which are often used in practice.

The models cited in this section all originate from training evaluation. Even though some of them are somewhat older and their beginnings go back to the 1950s, they are still relevant today. In particular, Kirkpatrick's 4-level model (first: 1959a) is still ubiquitous in training evaluation today.

Models from training evaluation are always interesting for the design of one's own evaluation project if the product to be evaluated is characterised by intentionality and teaching/learning objectives are perhaps even prepared with the help of learning objective taxonomies (e.g. Anderson, et al., 2009; Bloom & Krathwohl, 1956; Krathwohl, et al., 1966; Simpson, 1966, 1972). This is partly the case in out-of-school educational work with young people. The small selection of training evaluation models presented here, namely the models of Kirkpatrick (1959a, 1959b, 1960a, 1960b), Kaufman and Keller (1964), Phillips (1991) and Warr, Bird and Rackham (1970) show many similarities, but also striking differences that can be inspiring for one's own evaluation project.

The Kirkpatrick modell

Internationally, probably the best known model of training evaluation is Kirkpatrick's model. Kirkpatrick developed his model for the evaluation of trainings of the American aid organisation Heifer International at the end of the 1950s and published it for the first time in 1959 and 1960 with a series of four articles. Kirkpatrick structures training evaluation with four steps (Kirkpatrick, 1959a, 1959b, 1960a, 1960b) or levels (Kirkpatrick & Kirkpatrick, 2006): (1) response, (2) learning, (3) behaviour, and (4) outcomes.

At level **(1) Reaction,** the immediate reaction of the participants is evaluated, most likely expressed in terms such as pleasure or satisfaction. Typical questions or items are "I am completely satisfied with the training", "The venue was suitable" or "The trainer was competent", "The training was useful" or "The level of difficulty was just right".

Level **(2) Learning** is about what the participants have learned in the training. Depending on the content and objective of the training, in youth work this might be mainly procedural knowledge, certain skills, but also attitudes. Youth work in particular also works on attitudes (anti-racism training, democracy, human rights, etc.). In relation to STEM, an outstanding goal is to break down gender stereotypes and to broaden the personal spectrum of perception and creativity.

Level **(3) Behaviour** evaluates the extent to which the learning content of a training or what has been learned is actually reflected in the participants' later behaviour, i.e. what has been learned is actually applied. Evaluations at this level 3 show that this transfer of training is not automatically successful and that what has been learnt is only partially reflected in the participant's natural habitat. This transfer problem has been the subject of research over many years (e.g. Baldwin & Ford, 1988). An evaluation at level 3 should therefore be designed in such a way that the obstacles to transfer are also addressed. This broadens the focus, the individual participant is now seen as part of social (and socio-technical) systems. This is also where evaluation takes place, and in principle a wide range of methods and instruments are available for it: Questioning and observation of the respective participant, 360-degree feedback, logbook, etc.

At the level of **(4) Outcomes**, the training is evaluated in terms of the outcomes that are considered desirable by the organisation that initiated the participation. These will often be outcomes that are directly or indirectly of a financial nature. Kirkpatrick (1960b) sees here, for example, for an operational context, the reduction of costs, staff turnover, absenteeism, complaints; the increase of quality and quantity in production; and the improvement of the working atmosphere. This can be transferred to youth work organisations and institutions, including when it comes to developing young people's STEM skills and entrepreneurial attitudes. For example, a youth organisation such as the youth fire brigade or the youth in the German Federal Agency for Technical Relief (THW) will have an eminent interest in ensuring that training with its young people is also reflected in its organisational performance.

Kaufman's (and Keller's) model of learning evaluation

Roger Kaufman and John M. Keller published a 5-level evaluation model in 1994 in response to Kirkpatrick's very popular model. They take up critical voices on the practice of training evaluation at that time, which considered evaluation to be an undervalued endeavour that is often used inappropriately and sometimes has harmful consequences (Kelle and Möller, 1983). Kaufman and Keller (1994) suggest that there are at least three reasons for this: (1) the most commonly used definitions and models of evaluation are often too narrow; (2) the right questions are not asked for evaluation to answer; and (3) the relationship between goals and means is not made clear in the evaluation, planning and implementation processes (p. 371).

As (a) response to this, Kaufman and Keller (1994) broaden the focus of training evaluation to include resources that act as enablers of a learning process and the social consequences of the training activity/programme. Their model includes five stages or levels: Enabling (1a)/ Reaction (1b), Acquisition (2), Application (3), Organisational outputs (4) and Societal outcomes (5).

On the level (1a) Enabling, which is predominantly referred to as input in the receptions of the Kaufman model, all resources that flow into the teaching/learning process can be evaluated in principle. These are primarily the learning materials provided, the technical equipment, the qualification of the trainers, the premises. Looking further, however, the access requirements of the individual learners and the composition of the group are also important input factors or enablers.

At level **(1b) Reaction**, the focus is on the teaching/learning experience of the participants. Here, the acceptance and efficiency of the methods used, the resources and the process are recorded and evaluated. Following the first part, this second part is often also referred to as evaluation of the process.

At level **(2) Acquisition**, the acquisition of the learning content is recorded and evaluated. Kaufman and Keller speak here of mastery and competency, both of the individual learner and of a small group. It is therefore a question of whether and to what extent the participants have achieved the teaching/learning objectives. Evaluation at this level is comparable to Kirkpatrick's level 2.

At level **(3) Application**, it is recorded and evaluated whether and how well the participants apply what they have learned in their practice. Evaluation at this level is comparable to Kirkpatrick's level 3. In the Kaufman model, this level (together with level 2) is also called the micro-level.

At level **(4) Organisational Outputs**, the benefits to the organisation are recorded and evaluated. The evaluation is carried out from the perspective of the organisation that initiated participation in the training measure, on the slide of the results it wants. Evaluation at this level corresponds to level 4 of the Kirkpatrick model. Level 4 is also referred to as the macro-level in the Kaufman model.

Level **(5) Societal outcomes**, which has no equivalent in the Kirkpatrick model, and which is also referred to by the authors as the mega-level, opens the view to society as a whole. If training contributes to the effectiveness and efficiency of an organisation, the authors' evaluation model raises the question of whether and to what extent the strengthened organisation (and thus the training) makes a desirable contribution to the welfare of society. At level 5, questions about health, prosperity, environment, security, etc. are to be asked and answered; questions that have lost none of their relevance.

The Phillips ROI-model

Phillips (1996,1998) adds a fifth stage (or level) to the Kirkpatrick model, evaluating the ROI of a training intervention. The ROI or return on investment is a benchmark from financial management that indicates the relationship between profit and invested capital, thus an indicator of the profitability of an investment. Phillips transfers this indicator to human resource development or HRM, with his evaluation model he wants to show the value of a training investment in financial terms (Phillips, 1991). If the models of Kaufman (1994) and Phillips (1996,1998) are compared, the respective extension with the fifth level (and thus accentuation) could hardly be more different.

The five levels of the Phillips model are (1) reaction and planned action; (2) learning; (3) job application; (4) business results; and (5) return on investment/ROI.

Level **(1) Reaction**, like Kirkpatrick's Level 1, captures and assesses participant satisfaction with the training intervention. In addition, the Phillips ROI model asks what ideas or plans the participants have for applying what they have learned.

Level **(2) Learning** of the Phillips model essentially corresponds to that of the Kirkpatrick model. The focus is on learning gains, which are recorded and assessed on the basis of the teaching/learning objectives.

At level **(3) Application** in the workplace, the focus is - as in Kirkpatrick (1960a) and Kaufman and Keller (1994) - on behavioural changes in the workplace or, more generally, in the performance of tasks. What is evaluated is whether and to what extent what has been learned is actually transferred into practice and implemented there.

The evaluation at level **(4) Business results** of the Phillips Model has parallels to the Kirkpatrick Model, but the focus is narrower. At this level 4, the focus is on whether and to what extent participation in the training actually translates into business results for an organisation. This can be higher productivity, less time spent, lower costs, quality, higher customer satisfaction and much more. That this can also be interesting for organisations in youth welfare has already been explained in the discussion of the Kirkpatrick model.

At level (**5) Return on Investment/ROI**, the ultimate level of evaluation in Phillips' understanding, the monetary gain of training participation is compared with the costs. For him, determining the monetary cost-benefit ratio is the necessary step to complete an evaluation (Phillips, 1998).

CIRO (Context, Input, Reaction, Outcome) Modell

The CIRO model of Warr, Bird and Rackham (1970) closely follows a process of good planning and implementation of training and accordingly starts with evaluation even before the actual training activity takes place. It introduces a new level, which the authors call context evaluation. Otherwise, there are many parallels to the Kirkpatrick model, although some of them are structured in a new and different way. The CIRO model has four levels: (1) context, (2) input (3) response and (4) outcome.

At the **(1) Context** level, the quality of the analysis of the current (and also future) application situation, from which the training needs are then identified and the training goals determined, is evaluated. This extended focus takes into account the fact that serious errors at this planning level are reflected at the latest in the non transfer of what has been learned into practice. In this case, the CIRO model wants to identify the cause of the failure where it actually lies.

At level **(2) Input**, the quality of the implementation of the training measure, including its conception and planning, is evaluated. The CIRO model now focuses on whether and how information about possible training techniques and methods was collected, a selection made, the training design created and the training planned properly. Attention is also paid to whether and how the organisation's resources have been used in the best possible way to achieve the desired objectives. Then comes the actual input, namely the management and methodological-didactic implementation of the training. All these aspects are the focus of the input evaluation.

At the level of **(3) Reaction**, the participants' reactions are recorded and processed in order to improve the process. This includes the participants' views on the training they experienced as well as their suggestions for improvement.

At the **(4) Output** level, results with different scopes are analysed: the immediate results in the form of learning gains, the intermediate results at the point of transfer into practice and the ultimate results that the organisation can realise thanks to the training or the training programme. Thus, evaluation at this level is similar to evaluation at the three Kirkpatrick levels of learning, behaviour and results.

Application of existing evaluation models to non-formal and informal education

Non-formal and informal education or informal learning in a youth work organisation is a highly complex process, and the evaluation of it is as complex and demanding as the object of evaluation itself, given the multitude of influencing variables to be controlled. Accordingly, such evaluations require an expertise that is usually only found in highly specialised institutes. They make use of various models of educational programme evaluation, some of which have been used for many decades, especially in the United States of America, in local, regional and also national settings.

The literature on programme evaluation is almost impossible to survey, and an adequate presentation of the relevant models or approaches regularly fills at least an entire book (e.g. Stufflebeam & Coryn, 2014) or an entire website (e.g. BetterEvaluation, 2021). These are recommended reading when discussing one's own ideas on the evaluation approach in discussions with potential external contractors for an evaluation. At this point,

only three models or approaches should be briefly mentioned, without wishing to diminish the importance of the many other approaches such as those of Patton (1978) or Stufflebeam (1971). These are the models of Stake (1976), Fetterman (1994) and Cousins and Earl (2005), three models that are very much in line with the self-understanding and mission of youth work organisations.

Responsive Evaluation (Stake)

Robert Stake designed his approach of responsive evaluation as a contrast to an approach he called "higher-order evaluation", which was characterised by a formal determination of goals, standardised testing of student performance, value standards of educational staff and reports in the style of a scientific journal (1976, p. 19).

Stake contrasted this with his approach of responsive evaluation. His approach is characterised more by pedagogical questions than by objectives or hypotheses, methodologically it works with direct and indirect observation of programme participation, it addresses the plurality of value standards of different groups, not only those of the teaching staff, and it shows a continuous attention to the information needs of the different participants and stakeholders (1976, p. 19).

Responsive evaluation, in Stakes' understanding, "is not participatory evaluation, but it is organized partly around stakeholder concerns and it is not uncommon for responsive evaluation feedback to occur early and throughout the evaluation period" (Stake, 2003, p. 66).

Participatory Evaluation (Cousins und Earl)

Bradley Cousins and Lorna Earl (1992) understand their participatory evaluation approach as an extension of the stakeholder-based model, which places its emphasis on improving the use of evaluation through more intensive and broader participation of primary users in the applied research process. At the same time, this approach is embedded in the concept of professionalisation (here of teaching) and in the concept of the learning organisation or organisational learning (among others Argyris & Schön, 1978; Argyris, 1993; Huber, 1991; Levitt & March, 1988).

Empowerment Evaluation (Fetterman)

The empowerment evaluation approach (Fetterman, 1994) differs from the evaluation models mentioned so far in several ways. It is in a sense programmatic and places its intention of enabling improvement and promoting people's self-determination centrally. To this end, "any means is welcome", i.e. also other evaluation concepts, a broad spectrum of evaluation techniques and the cooperatively developed evaluation results. Consistently, Fetterman also assigns new roles to the participants and the evaluators. The role of the evaluators is to provide those involved in the education programme with the necessary procedures and tools to assess the planning, implementation and self-evaluation of their education programme. Ideally, they empower them in such a way that they are able to continue evaluating and improving their education programme on their own, even beyond the evaluation.

The realisation of an empowerment evaluation is very flexible and possible with all conceivable instruments. 10 principles (Fetterman, 2005), a very simple three-step structure (Fettermann, 2001), which was later refined into a ten-step approach (Chinman, Imm and Wandersman, 2004), and a (non-finalised) fund of instruments help to realise empowerment evaluation successfully.

Conclusion

The aim of this paper was to provide some basic guidance on the essential aspects of evaluation for educational staff in youth work who are faced with the decision to evaluate a training, an educational project or even an educational programme. To this end, the paper first discussed the ethical challenges of an evaluation and the importance of clarifying the goals.The paper then presented the advantages and

disadvantages of external and internal evaluation, followed by a discussion of the two forms of evaluation, "summative" and/or "formative". The practical planning of an evaluation was then in focus, complemented with a presentation of four evaluation models applicable in non-formal education and references to three programme evaluation models that are very much in line with the self-image and mission of youth work organisations.

References

Abele, C., & Erdmann-Kutnevic, S. (2016). *Methodenheft zur Selbstevaluation von internationalen Jugendbegegnungen im Rahmen des Förderprogramms EUROPEANS FOR PEACE der Stiftung »Erinnerung, Verantwortung und Zukunft« (EVZ)*. Retrieved from https://www.stiftung-evz.de/fileadmin/user_upload/EVZ_Uploads/Handlungsfelder/Handeln_fuer_Menschenrechte/Europeans_for_Peace/Handreichung_Selbstevaluation.pdf

Argyris, C. (1993). *Knowledge for action: A guide to overcoming barriers to organizational change* (1st ed.). *The Jossey-Bass management series*. San Francisco, Calif.: Jossey-Bass Publishers. Retrieved from http://www.loc.gov/catdir/bios/wiley044/92042861.html

Argyris, C., & Schön, D. A. (1978). *Organizational learning*. *Organization development deries*. Reading, Mass: Addison-Wesley Pub. Co.

BetterEvaluation (May 2013). BetterEvaluation: Sharing information to improve evaluation. Homepage. Retrieved from https://www.betterevaluation.org/en

Bloom, B. S., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals*. Handbook I: Cognitive domain. New York: Longman.

Chinman, M., Pamela Imm, & Abraham Wandersman (2004). *Getting To Outcomes 2004: Promoting Accountability Through Methods and Tools for Planning, Implementation, and Evaluation*. Santa Monica, CA: RAND Corporation. https://doi.org/10.7249/TR101

Council of Europe, & Conseil de l'Europe (September 2018). *Self-assessment tool for youth policy*. Brussels. Retrieved from https://rm.coe.int/self-assessment-tool-for-youth-policy-english/16808d76c5

Council of Europe, & European Commission (2021). Glossary on youth: Training. Retrieved from https://pjp-eu.coe.int/en/web/youth-partnership/glossary

Council of Europe Youth Work (2021). Help: self-assessment: The Council of Europe Youth Work Portfolio. Retrieved from https://www.coe.int/en/web/youth-portfolio/help-self-assessment

Cousins, J. B., & Earl, L. M. (Eds.) (2005). *Teachers' Library*. *Participatory Evaluation In Education: Studies Of Evaluation Use And Organizational Learning*. London, Washington, D.C.: Falmer Press. Retrieved from http://gbv.eblib.com/patron/FullRecord.aspx?p=201009

Cousins, J. B., & Earl, L. M. (2005). The Case for Participatory Evaluation: Theory, Research, Practice. In J. B. Cousins & L. M. Earl (Eds.), *Teachers' Library. Participatory Evaluation In Education: Studies Of Evaluation Use And Organizational Learning* (pp. 3–18). London, Washington, D.C.: Falmer Press.

Cronbach, L. J. (1963). Course improvement through evaluation. *Teachers College Record*, *64*(8), 672–683.

Eval-Wiki: Glossar der Evaluation (2020, November 5). Evaluationsplan. Retrieved from https://eval-wiki.org/glossar/Evaluationsplan

Fetterman, D. M. (1994). Empowerment Evaluation. *Evaluation Practice*, *15*(1), 1–15. https://doi.org/10.1177/109821409401500101

Fetterman, D. M. (2005). A Window into the Heart and Soul of Empowerment Evaluation: Looking through the Lens of Empowerment Evaluation Principles. In D. M. Fetterman & A. Wandersman (Eds.), *Empowerment evaluation principles in practice* (1-26). New York: Guilford Press.

Fetterman, D. M., & Wandersman, A. (Eds.) (2005). *Empowerment evaluation principles in practice*. New York: Guilford Press.

Guba, E. G., & Lincoln, Y. S. (2001). *Guidelines and Checklist for Constructivist (aka Fourth-Generation)*. Retrieved from Western Michigan University website: https://wmich.edu/sites/default/files/attachments/u350/2018/const-eval-guba%26lincoln.pdf

Hart, R. A. (1992). *Children's participation: From tokenism to citizenship*. *Innocenti essays: no. 4*. Florence Italy: UNICEF International Child Development Centre. Retrieved from https://www.unicef-irc.org/publications/pdf/childrens_participation.pdf

Huber, G. P. (1991). Organizational Learning: The Contributing Processes and the Literatures. *Organization Science*, *2*(1), 88–115. Retrieved from http://www.jstor.org/stable/2634941

Kaufman, R., & Keller, J. M. (1994). Levels of evaluation: Beyond Kirkpatrick. *Human Resource Development Quarterly*, *5*(4), 371–380. https://doi.org/10.1002/HRDQ.3920050408

Kellaghan, T., & Stufflebeam, D. L. (Eds.) (2003). *Kluwer International Handbooks of Education: Vol. 9*. *International Handbook of Educational Evaluation*. Dordrecht: Springer. https://doi.org/10.1007/978-94-010-0309-4

Kelle, J. M., & Möller, J. M. (1983). Evaluation of Training: Much Lauded, Seldom Applauded. *IFAC Proceedings Volumes*, *16*(6), 201–209. https://doi.org/10.1016/S1474-6670(17)64365-9

Kirkpatrick, D. L. (1959). Techniques for Evaluation Training Programs: Part 1 - Reaction. *Journal of the American Society of Training Directors*, *13*(11), 21–26.

Kirkpatrick, D. L. (1959). Techniques for Evaluation Training Programs: Part 2 - Learning. *Journal of the American Society of Training Directors*, *13*(12), 21–26.

Kirkpatrick, D. L. (1960). Techniques for Evaluation Training Programs: Part 3 - Behavior. *Journal of the American Society of Training Directors*, *14*(1), 13–18.

Kirkpatrick, D. L. (1960). Techniques for Evaluation Training Programs: Part 4 - Results. *Journal of the American Society of Training Directors*, *14*(2), 28–32.

Kirkpatrick, D. L., & Kirkpatrick, J. D. (2006). *Evaluating training programs: The four levels* (3. ed.). San Francisco, Calif., London: Berrett-Koehler; McGraw-Hill. Retrieved from http://www.gbv.de/dms/bsz/toc/bsz278284698inh.pdf

König, J. (2007). *Einführung in die Selbstevaluation: Ein Leitfaden zur Bewertung der Praxis Sozialer Arbeit* (2., neu überarbeitete Auflage). Freiburg im Breisgau: Lambertus-Verlag. Retrieved from https://www.lambertus.de/assets/adb/94/94d43588099c1578.pdf

Krathwohl, D. R. (2002). A Revision of Bloom's Taxonomy: An Overview. *Theory into Practice*, *41*(4), 212-218. Retrieved from https://www.depauw.edu/files/resources/krathwohl.pdf

Krathwohl, D. R., Bloom, B. S., & Masia, B. B. (1956). *Taxonomy of Educational Objectives: The Classification of Educational Goals*. Handbook II: Affective Domain. New York: David McKay Company.

Levitt, B., & March, J. G. (14). Organizational Learning. *Annual Review of Sociology*, *1988*, 319–340. Retrieved from http://sjbae.pbworks.com/f/levitt_march_1988.pdf

Patton, M. Q. (1978). *Utilization-focused evaluation*. Beverly Hills, Calif.: SAGE.

Phillips, J. J. (1996). How Much Is the Training Worth? *Training and Development*, *50*(4), 20–24.

Phillips, J. J. (1998). The Return-on-Investment (ROI) Process: Issues and Trends. *Educational Technology*, *38*(4), 7–14.

Pollitt, C. (1998). Evaluation in Europe. *Evaluation*, *4*(2), 214–224. https://doi.org/10.1177/13563899822208554

SALTO Training and Cooperation Resource Centre (2021). YOCOMO self-assessment tool for youth workers. Retrieved from https://satool.salto-youth.net/

Scriven, M. [Michael] (1967). The Methodology of Evaluation. In R. W. Tyler, R. M. Gagné, & M. Scriven (Eds.), *Monograph series on curriculum evaluation: Vol. 1. Perspectives of Curriculum Evaluation* (pp. 39–83). Chicago: MacNally. Retrieved from https://files.eric.ed.gov/fulltext/ED030948.pdf

Scriven, M. [Michael] (1981). *Evaluation Thesaurus* (3rd edition). Retrieved from https://files.eric.ed.gov/fulltext/ED214952.pdf

Simpson, E. J. (1966). *The Classification of Educational Objectives, Psychomotor Domain*. Report Resumes. Retrieved from University of Illinois website: https://files.eric.ed.gov/fulltext/ED010368.pdf

Simpson, E. J. (1972). *The Classification of Educational Objectives in the Psychomotor Domain*. Washington, DC: Gryphon House.

Stake, R. (2003). Responsive Evaluation. In T. Kellaghan & D. L. Stufflebeam (Eds.), *Kluwer International Handbooks of Education: Vol. 9. International Handbook of Educational Evaluation* (pp. 63–68). Dordrecht: Springer. https://doi.org/10.1007/978-94-010-0309-4_5

Stake, R. E. (1976). A theoretical statement of responsive evaluation. *Studies in Educational Evaluation*, *2*(1), 19–22. https://doi.org/10.1016/0191-491X(76)90004-3

Stockmann, R. (2004). *Was ist eine gute Evaluation? Einführung zu Funktionen und Methoden von Evaluationsverfahren* (CEval Arbeitspapiere No. 9). Saarbrücken:.

Stufflebeam, D. L. (1971). The Relevance of the CIPP Evaluation Model for Educational Accountability. *Journal of Research and Development in Education*. Retrieved from https://files.eric.ed.gov/fulltext/ED062385.pdf

Stufflebeam, D. L., & Coryn, C. L. S. (2014). *Evaluation Theory, Models, and Applications* (2nd edition). *Research methods for the social sciences: Vol. 50*. San Francisco CA: Jossey-Bass.

Tyler, R. W., Gagné, R. M., & Scriven, M. [M.] (Eds.) (1967). *Monograph series on curriculum evaluation: Vol. 1*. *Perspectives of Curriculum Evaluation*. Chicago: MacNally. Retrieved from https://files.eric.ed.gov/fulltext/ED030948.pdf

United Nations Children's Fund (UNICEF): Office of Research - Innocenti (September 2013). *Ethical Research Involving Children*. Florence, Italy. Retrieved from https://childethics.com/wp-content/uploads/2013/10/ERIC-compendium-approved-digital-web.pdf

Warr, P., Bird, M., & Rackham, N. (1970). *Evaluation of management training: A practical framework, with cases, for evaluating training needs and results*. London: Gower P.

Yarbrough, D. B., & Shula, Lyn M., Hopson, Rodney K., Caruthers, Flora A. (2011). *The program evaluation standards: A guide for evaluators and evaluation users* (3. ed.). Thousand Oaks: SAGE Publ. Retrieved from https://evaluationstandards.org/wp-content/uploads/2019/08/Yarbrough-3e-Rev-Errata-Page.pdf

YouthLink Scotland (2021). Self-evaluation and improvement. Retrieved from https://www.youthlinkscotland.org/policy-research/outcomes-and-skills/youth-work-outcomes/about-the-youth-work-outcomes/self-evaluation-and-improvement/