

# Monitoring and measuring the results of interventions related to technical and vocational education and training and the labour market

A guideline for practitioners





# **Monitoring and measuring the results of interventions related to technical and vocational education and training and the labour market**

A guideline for practitioners

Sector programme 'Technical and vocational education  
and training'

Dr. Wolfgang Meyer  
Stefan Thomas



*Dear Reader,*

Technical and vocational education and training (TVET) has been an important area of intervention in German development policy for many years. As part of integrated education systems TVET contributes significantly to improving the living conditions in our partner countries, both at an individual and a societal level.

The Federal Ministry for Economic Cooperation and Development regards education and TVET as pivotal success factors for combating poverty and for sustainable economic and social development. Our goal is lifelong education for all. This is why we will boost our investments in education up to 2013: we will increase the number of partner countries in which our development policy focuses on education and by 2013 we will at least double the funds for education in Africa. This commitment to education is also an appeal to all those responsible for ensuring that investments in the education systems of our partner countries are as effective and efficient as possible. In this respect we seek constant improvement and have accordingly chosen 'Results' as our leitmotif for 2012. This guideline is perfectly timed to support us in this endeavour. It outlines the methodology and instruments that can be used to measure the results of our work. Monitoring and measuring results is not just a basis for transparent public information about our work but is also indispensable if we are to gain knowledge on how investment in education in our partner countries should be conceived in the years to come. At this point I wish to thank the authors and all those who were involved in producing this guideline and hope that reading it may provide useful insights and inspiration for putting it into effect in your projects and programmes.



*Roland Lindenthal*  
*Federal Ministry for Economic Cooperation and Development*  
*Head of Education Division*

## Table of Figures

Figure 1: Seven steps for establishing a results-based monitoring .....	6
Figure 2: Terms used in results chains .....	9
Example: Developing a results chain .....	10
Figure 3: Results chain and key questions for developing an intervention .....	11
Figure 4: Scaling up achieved results (diffusion/widespread impact) .....	13
Figure 5: Results chain and core monitoring questions .....	20
Figure 6: Monitoring plan for the case study curricula development .....	22
Figure 7: Results chain and attribution of changes .....	29
Figure 8: Simple before/after comparison without control group/comparison group .....	30
Figure 9: Example for an experiment .....	31
Figure 10: Example for a quasi-experiment .....	32
Figure 11: Example for a cross-sectional comparison .....	33

# Table of Contents

<b>1. Introduction</b>	<b>6</b>
<b>2. Results and results chains</b>	<b>8</b>
2.1 What are results?	8
2.2 How to word results hypotheses	8
2.3 What is a results chain?	9
2.4 What is the point of results chains?	11
2.5 Results chains checklist	14
<b>3. Indicators</b>	<b>15</b>
3.1 What are indicators?	15
3.2 How should indicators be drawn up?	15
3.3 How to assess the quality of indicators	17
3.4 What are results-based indicators?	17
3.5 What are process indicators?	18
3.6 Using indicators for results-based monitoring	19
3.7 Indicator checklist	24
<b>4. Principles of measurement</b>	<b>25</b>
4.1 How to assess the quality of measurements?	25
4.2 Which particular challenges arise when measuring the results of TVET?	26
4.3 Measurement of indicators checklist	28
<b>5. Measuring interrelation of results</b>	<b>29</b>
5.1 Methods to verify the interrelation of results	30
5.1.1 Experimental design	31
5.1.2 Quasi-experimental design	32
5.1.3 Cross-sectional comparison	33
5.1.4 Qualitative design	34
5.2 Verifying the interrelation of results using curriculum development as an example	35
5.3 Further examples for the application of strategies and data collection methods to verify the interrelation of results	42
5.3.1 Optimizing the organisational and management structure of vocational schools	42
5.3.2 Optimizing career guidance, employment services and training in employment agencies	44
5.3.3 Optimizing the entire national TVET system	46
<b>6. Final remarks: 10 tips for successful results-based monitoring</b>	<b>48</b>
<b>7. Sources of reference and further reading</b>	<b>49</b>
<b>8. Glossary</b>	<b>52</b>

## 1. Introduction

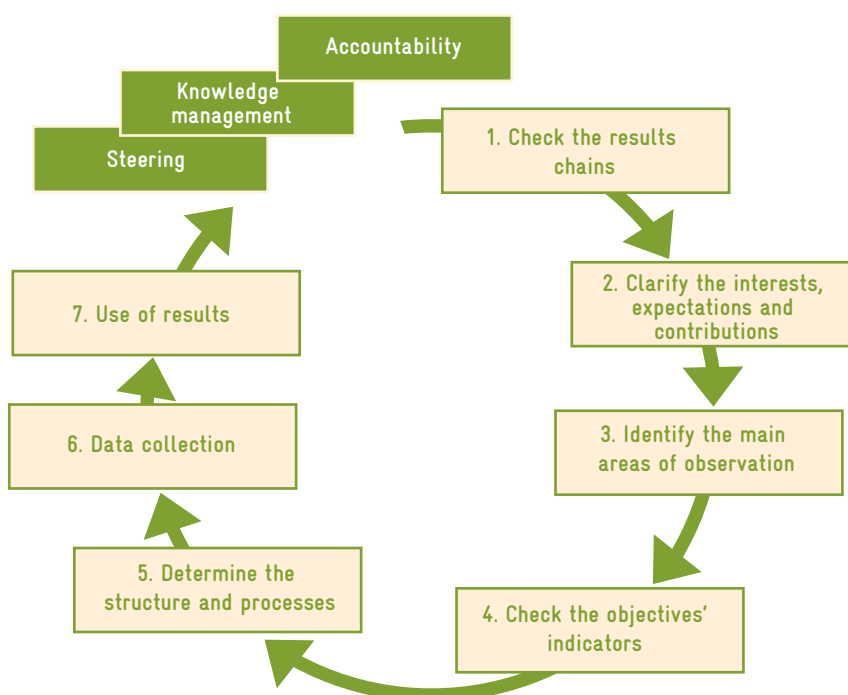
Monitoring and measuring results are imperatives for practitioners in international development cooperation, since they are the only means to provide evidence of the results and effectiveness of development work in partner countries. When GIZ introduced its framework for contracts and cooperation (AURA, by its German acronym), it geared all its development policy initiatives towards achieving results. This orientation towards results differs substantially from the previous procedure, both in terms of the underlying results chains and the indicators used to measure them (cf. GTZ 2006).

Planning, monitoring and evaluation are used to reflect the results of development measures and programmes. Results-based monitoring in particular is a fundamental part of project management. By continuously checking the effects of an intervention it is possible to identify positive and negative developments early enough to address them. It makes clear which measures are working and which are not achieving results or not the desired results. Monitoring is also the basis for sound accountability.

Successful results-based monitoring requires specifying the objectives of the intervention and defining appropriate indicators to measure their achievement. It is the latter which poses a particular challenge for the persons in charge of the project. Not only is it difficult to define and measure results, it is also often not easy to establish which positive changes can be attributed to the measures undertaken by the project and which have been caused by other factors.

This guideline builds on the GIZ's general 'Results-based Monitoring – Guidelines for Technical Cooperation Projects and Programmes'. These general guidelines describe seven steps for results-based monitoring. The following figure shows these seven steps<sup>1</sup>:

**Figure 1: Seven steps for establishing a results-based monitoring**



1 For a detailed description of all steps please refer to the results-based monitoring guidelines themselves (GTZ 2008).

For results-based monitoring to succeed it is essential to follow these seven steps. The structure of this guideline follows this rationale. After clarifying some definitions and the purpose of results and results chains in the first part, the following chapter takes a closer look at the purpose of indicators and how to identify and measure them. The last part of the guideline deals in detail with how to measure the correlation between observed results and the intervention ('attribution problem').

For a better understanding of the core elements of results-based monitoring and measurement, these are illustrated by one example (curriculum development) throughout the text. Of course such an example cannot entirely reflect the complex reality of technical and vocational education and training (TVET). However the principles and methods explained here can be easily transferred to other diverse contexts of development projects.

The glossary in the annex explains important terms related to monitoring and measuring results.

## 2. Results and results chains

### 2.1 What are results?

Results are *changes* to a certain condition due to the occurrence of an *intervention* (or the lack of it). For instance education measures are based on the assumption that training (intervention) will lead to a learning outcome (increased or new skills) for the participants of the training. Accordingly, if a training measure did not lead to increased or new skills it would produce no results.

However, the fact that a change has occurred is not enough to draw any conclusions regarding the *results of a certain measure*. The learning outcome of the participants can also have other causes (for instance, they may have participated in other courses, read literature related to the same topic or discussed the issue with experts), i.e. even if we can prove there has been a learning outcome, it may still be the case that the training measure produced no results.

Of course the people involved in planning and conducting training assume that it will lead to a learning outcome. This means that – consciously or unconsciously – a causal relationship is established between a cause (the training) and its *intended result* (increase of skills). This causal relationship can be analysed in order to verify its validity. These causal assumptions are also called *results hypotheses*.

### 2.2 How to word results hypotheses

Hypotheses establish a connection between two facts and can be expressed in various ways. They can be worded as ‘if-then-statements’ (‘If the colleges use new curricula then the graduates will be more employable’) or as ‘the-the-statements’ (‘The more people have the necessary skills to develop curricula, the more needs-oriented curricula will be developed’).

Important elements which should be part of any hypotheses development are:

- a) The cause (‘if’-component, ‘independent variable’): What is supposed to provoke a reaction?
- b) The effect or result (‘then’-component, ‘dependent variable’): Which reaction is expected to follow?
- c) The connection (‘if-then’, ‘the-the’): Which correlation is assumed? There is a positive correlation if increasing the input leads to a rise in results. We speak of a negative correlation when increasing the input leads to a decrease in results.
- d) The kind of correlation: How is the result expected to evolve in relation to changes in its origins? Usually a linear correlation is assumed (i.e. the cause and its results vary in equal degree), but any other mathematical form (e.g. an exponential increase or the oscillating fluctuations of a sinus curve) may be applicable and can be expressed as part of a hypothesis.

## 2.3 What is a results chain?

The planning rationale of a project or programme is usually expressed as a sequence of consecutive causal connections, i.e. one single cause-and-effect connection (results hypothesis) turns into a *results chain*, where the causes are (at least partly) the results of previous activities and the achieved outcomes at least partly contribute to further impacts.

According to this concept, a results chain consists of six core elements<sup>2</sup>, which are causally connected to each other by means of results hypotheses. Following the planning rationale, the *activities* of a programme or project are expected to produce certain *outputs*. The next step is the expected *use of outputs* by the intended target groups and in accordance with the intentions of the providers of the outputs (these are the partner organisations participating in the programme or project). It is only through this use of the outputs that results can be achieved. Here it is important to differentiate between results arising directly from the use of outputs (*outcome* or *direct benefit*) and more far reaching results (*impact* or *indirect benefit*), which can only be indirectly attributed to the activities of the project or programme. The impacts contribute to the achievement of the highly aggregated results at development policy level (Millennium Development Goals, MDGs).

The following table shows the terms normally used in results analysis.

**Figure 2: Terms used in results chains**

German	English
Indirekte Wirkung	Impact/Indirect Benefit
Direkte Wirkung	Outcome/Direct Benefit
Nutzung der Leistung	Use of Output
Leistung	Output
Aktivitäten	Activities
Inputs	Inputs

---

<sup>2</sup> Results-based Monitoring – Guidelines for Technical Cooperation Projects and Programmes (GTZ 2008).

### **Example: Developing a results chain**

A TVET project aims at improving the quality of training, so that more people can have better-paid employment and thus earn a higher income as well as contribute to increasing the economic power of the country due to their improved productivity. While planning the project, the following results chain was drawn up:

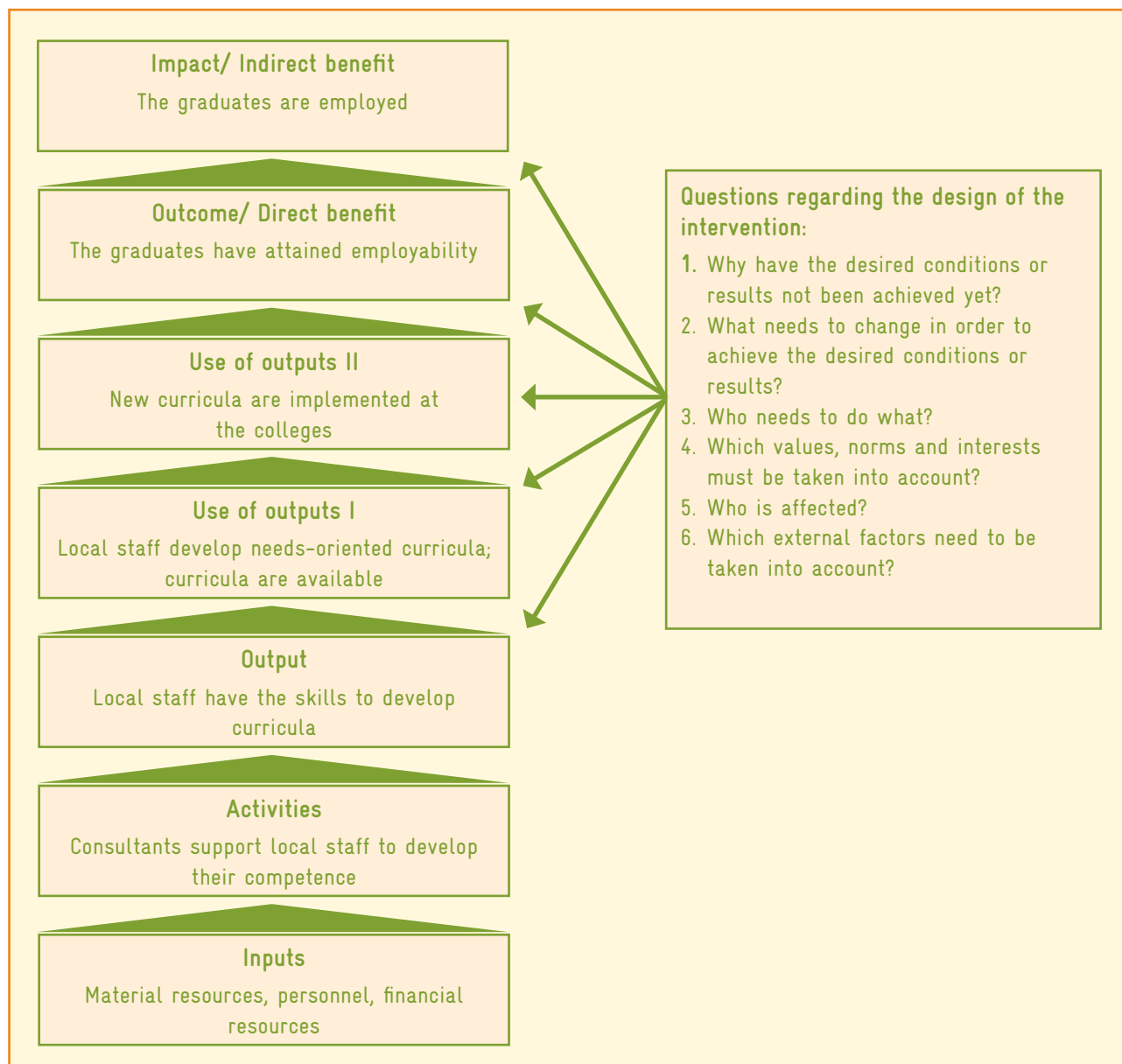
#### *Results hypotheses:*

The expected results can be expressed as the following hypotheses:

1. If the project activities are carried out, then the local staff's skills for developing appropriate curricula will have improved.
2. If curricula are developed by the people who have received training, then they are tailored to the needs of the trainees and the labour market.
3. If the needs-oriented curricula are used in TVET schools, then the employability of the trainees is improved, i.e. their chances of finding employment that is in line with their level of training increase when compared with trainees who have not undergone training based on needs-oriented curricula.
4. The more trainees receive reformed, labour-market-oriented training, the more trainees find employment that corresponds to their level of training.

If employment that is in line with the trainees' level of training leads to a positive income effect that matches the increase in productivity, then the programme eventually contributes to reducing poverty and therefore to the achievement of Millennium Development Goal 1.

**Figure 3: Results chain and key questions for developing an intervention**



## 2.4 What is the point of results chains?

Results chains connect the changes expected to occur at different levels in the course of an intervention. By splitting the information up into verifiable results hypotheses the results chain helps to answer the questions about why the desired changes have not taken place yet and what needs to happen to change this. This is important *support for programme or project management*.

In the case of our example, it may turn out that although the training measures do lead to an increase in competence (Hypothesis 1), they do not lead to the development of needs-oriented curricula (Hypothesis 2), for instance because the members of staff who receive training are not authorised to make decisions about the curriculum content. As soon as this is recognised, the intervention can undertake the necessary measures (e.g. additional awareness raising of the management staff of the schools concerned) to improve the results.

Given that results hypotheses are always assumptions which may prove to be wrong during the course of the intervention, it is important to *check them regularly*. This is especially the case in development cooperation projects or programmes, as the context in which the intervention operates is subject to continuous development changes – meaning that even tried and tested instruments may lead to different results over time.

Developing and checking a results chain does not only concern the staff of the project or programme. At the intervention's macro level, it is particularly important to ensure the support of the national government, administration, academia, etc. (for instance in terms of supplying data, statistics, research results etc.) (cf. chapter 5). Therefore it is essential to make sure *partner organisations participate* but also that *appropriate institutions in the partner country provide support for results-based monitoring*.

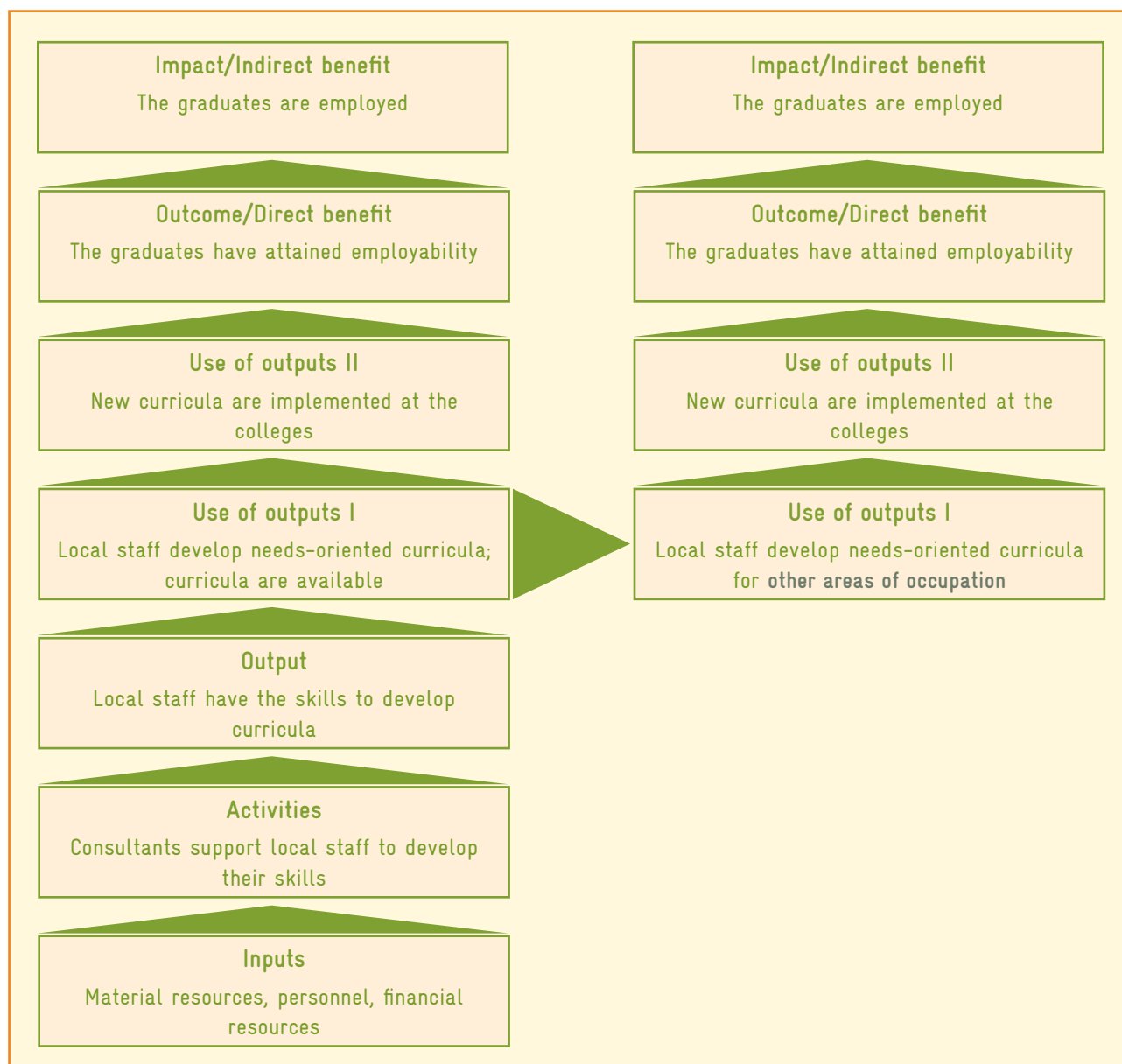
This is particularly important in areas which lie beyond the direct responsibility of the project and which contribute to turning the project's *outcomes* into *impacts* from the perspective of development policy. *Scaling up the results* achieved (so-called dissemination or widespread impact) across the entire national context or sector and ensuring the *durability of these results* (so-called sustainability) are crucial aspects to consider. This is the only way development cooperation interventions can achieve their intended profound changes in the system and long-term success in terms of development objectives.

It is the partner countries and their institutions that are solely responsible for managing these processes and for using appropriate monitoring and evaluation instruments to measure progress. Ideally, the capacity development measures undertaken by the project as well as its indicators provide a solid basis, which local institutions can then take the responsibility to expand.

### **Integrating widespread impact into the results chain**

Looking at our example, it is important to ensure that the capacities developed by the project – the ability and preparedness of local staff to develop needs-oriented curricula – are used on a sustainable and wide-spread scale. This might entail local staff *not only* applying their new capacities within an *area initially prioritised* by the project (e.g. development of curricula for information and communication technologies) *but expanding* their activities *to include other areas of occupation* (e.g. development of curricula for the construction sector). These assumptions can also be expressed in terms of results hypotheses and thus be integrated into the results chain. This way, it is possible to expand and complement the results chain to include other elements as needed. The basic structure of the single elements (inputs, activities, outputs, use of outputs, outcome, and impact) remains the same (see Figure 4).

**Figure 4: Scaling up achieved results (diffusion/widespread impact)**



Results tend to materialise with a certain *time lag*, i.e. not immediately and parallel to the intervention but after a more or less long period of time afterwards. This poses a particular challenge in the case of *measures geared towards changing entire systems*, which usually take several years to produce the desired results at the target group level. Here *process indicators* (see chapter 3.5) can provide helpful information about the current and future development of results within the context of results-based monitoring. They do not necessarily express achieved results but point to certain prior changes or developments within the process.

## 2.5 Results chains checklist

Requirements of results chains	
Are the results chains logical?	<input type="checkbox"/>
Does the results chain contain all relevant changes/results?	<input type="checkbox"/>
Are the results chains intelligible for outsiders?	<input type="checkbox"/>
Are the results chains detailed enough for strategic management?	<input type="checkbox"/>

Using results chains for management	
Does the intervention have a comprehensive results chain showing the rationale of the entire programme?	<input type="checkbox"/>
Does the intervention have detailed result chains for all its main lines of action?	<input type="checkbox"/>
Do all project staff know the results chains?	<input type="checkbox"/>
Are the results chains used as the basis for strategic discussions?	<input type="checkbox"/>
Are the results chains adapted at least once a year?	<input type="checkbox"/>

## 3. Indicators

### 3.1 What are indicators?

According to GIZ's definition, indicators are 'parameters, which are used to express a certain, often complex condition, which cannot be measured directly' (GTZ 2004: 89). Literally, an 'indicator' is a means of making something known that might otherwise not be visible to the observer (cf. Meyer, 2007, p. 195 et seq.). In development cooperation, as in other areas, this 'something' often cannot be measured directly but is a *theoretical construct*<sup>3</sup>: for instance 'language competence' can only be determined and reflected on a uniform scale by means of tests requiring considerable effort (e.g. the TOEFL test). Even though the TOEFL test – like most other tests – does not directly measure the language competence but uses a series of language-related statements as *indicators*.

GIZ's definition also points to an additional, extremely positive property of indicators: they can *simplify barely comprehensible issues* making them readily understandable. The gross national product, for instance, is used to express the value of an entire economy thus making it possible to *compare* it to other economies. The amount of information required and the effort needed to calculate the gross national product are enormous – the result, however, is a single figure expressed in terms of a unit (the local currency) which can be easily understood by a layperson.

'Good' indicators thereby produce 'clear' results in a twofold sense: they allow us to establish exact findings on complex issues and are not only easy to understand but also to interpret.

### 3.2 How should indicators be drawn up?

The requirements for drawing up indicators are largely the same as those proposed for the definition of objectives in project management literature<sup>4</sup>. This lies in the nature of indicators themselves which usually refer to the objectives of the project or programme and are expected to reflect their achievement (or the results produced by the achievement of the objectives).

---

3 A theoretical construct is a complex (social) phenomenon containing various dimensions. The terms 'happiness', 'satisfaction', 'peace', 'democracy', 'quality of teaching', etc. are some examples. In our example this particularly applies to the term 'employability'.

4 Cf. for instance PMA 2008: 27; similar and much more detailed in the PMBOK-Guide, the international standards for programme management of the Programme Management Institute (PMI 2008). The following source (in German) is available online for free: <http://www.akademie.de/thema/projektmanagement>

### **An indicator should therefore**

- ▶ describe an object (what does the indicator refer to, what should it reflect).
- ▶ denote the content of the measurement (what should be measured).
- ▶ define the scale of the measurement (what do the measured values mean).
- ▶ define the measurement location (where should measuring take place).
- ▶ and determine the timing for measurement (when or at which intervals something should be measured and which time period should the resulting statements refer to)

In the case of our example the objective at impact level is that the graduates of the reformed training programmes find employment.

The ‘graduates’ rate of employment’ will only be a useful indicator if its definition fulfils the five requirements explained earlier in this section. This measure can – just like any other measure – reflect a variety of constructs and relate to these constructs in different ways. The rate of employment might be used as a means to express poverty reduction in the families of the target region, identify the usefulness of the training, indicate the local labour market conditions or show the willingness of local enterprises to employ certain groups of people.

In terms of the added value of a certain kind of training in comparison with other training programmes, this indicator could also be defined as follows:

‘The added value of a certain training programme is reflected by the employment rate (as a percentage) of qualified workers who are employed for the first time after completing a training programme in the area of occupation X. Measurement takes place yearly half a year after the different vocational schools have completed their courses and in all enterprises in the target region y. The employment rate of the project’s participants should be 20 percent above the average employment rate.’

The first sentence is the actual *definition of the indicator*, which should be as easily understandable as possible. Above all, it should express the indicator’s measuring goal and its unit. The second sentence supplements the measuring rules by means of a *technical explanation*, which is mandatory for its replication by other people but not necessarily for its interpretation in terms of its content. This is why such explanations are usually not part of the main body of a report but are explained in an annex on methodology. Finally, the third and last sentence of the example refers to the use of the information provided by the indicator. *Assigning a numerical value* to the *indicator* sets the target to be reached and therefore makes it possible to conduct target-performance analyses.

### **According to GIZ’s rationale for project proposals the indicator could be specified as follows:**

‘The employment rate of the graduates of reformed training programmes in the area of occupation X is 20 percent above the average employment rate.’ (Tracer study to be conducted six months after course completion in the region y).

### 3.3 How to assess the quality of indicators

The *usefulness of an indicator* lies in being able to compare different results with each other. In the case of monitoring, the purpose is mainly to conduct target-performance analyses.

In project planning indicators are drawn up to reflect the goals of a certain intervention and are expected to ‘state how the intended change can be observed or measured’ (cf. The world of Words at GTZ 2007:83). The underlying idea, therefore, is that when planning a TVET intervention, a target status should be agreed upon with the local partners to be reached in certain phases and verified by means of indicators.

A baseline study conducted at the beginning of the intervention offers the opportunity to compare the actual situation with the target status and therefore provides information on the state of development.<sup>5</sup> A list of indicators that reflect the objectives of the project can be used to regularly measure (*actual status*), control and document the degree of achievement of the project’s objectives. Since AURA was introduced, the target values of objectives are expressed as results.

The quality of indicators can be assessed according to a variety of criteria. A relatively common pattern is the so-called SMART-criterion<sup>6</sup> which uses the following five elements to assess the quality of indicators:

#### **The five SMART-criteria:**

- ▶ Specific: an indicator must measure exactly one area and always render the same result despite changes in other adjacent areas.
- ▶ Measurable: it should be possible to measure an indicator with sufficient precision and with a reasonable effort.
- ▶ Achievable: an indicator should provide sufficiently precise results that are useful for the pursued purpose.
- ▶ Relevant: an indicator should provide results that are sufficiently exact for the (sub-) area it is set to describe.
- ▶ Timebound: finally, an indicator should be able to provide the results in time for their utilisation.

### 3.4 What are results-based indicators?

What is special about results-based indicators is that they come ‘at the end of a complex and complicated cause-and-effect chain’ (Haldemann 2009). They show *outcomes* and *impacts* – instead of ‘inputs’ or *outputs*. In the present example on ‘curricula development’ we need to define indicators that show whether the graduates of reformed training programmes have attained employability (outcome) or whether they have found employment (impact). If the objectives of an intervention are defined as results, then the indicators at outcome level are at the same time objective indicators – as is the case with

---

5 Cf. ‘Baseline Studies – A Guide to Planning and Conducting Studies, and to Evaluating and Using Results’ (GTZ 2010).

6 There are different and sometimes contradictory interpretations of the criteria behind the acronym ‘SMART’. Also, the SMART-criterion overlooks several aspects important for assessment (such as the social component of generating and using indicators). Even so, the SMART-criterion provides good initial support for guiding the process of indicator assessment.

project proposals to the German Federal Ministry for Economic Cooperation and Development (BMZ).

The general *requirements for drawing up* indicators also apply to results-based indicators. Due to the causal relationship between causes and results, however, special attention needs to be paid to the development of results hypotheses, which in turn will have consequences for the development of the indicators. This was already mentioned above (see also chapter 5).

*Establishing results-based indicators* is a discussion process between all those involved on what is doable and it sharpens the eye for realistic expectations of results. Only by discussing which changes are expected to happen can the *measuring goal of the indicators* be defined. This is why results-based indicators are always the product of a specific project or programme.

### 3.5 What are process indicators?

The task of drawing up indicators cannot be limited to results-based indicators alone. Process indicators that reflect all the other elements of the results chain (inputs, activities, outputs, use of outputs) are also involved. It is particularly important for project management to be able to assess the project's degree of progress at all levels by means of target-performance analyses conducted as part of regular monitoring (see chapter 4.1).

It is quite easy to check parameters such as the amount of financial resources available or the amount of training conducted. In most cases it is still very helpful to use indicators at the level of *inputs* and *activities*. This may be necessary in order to calculate further measures (for instance the ratios used to assess efficiency such as 'input-output', 'cost-benefit', 'cost-results', etc.). At the level of activities, our example reads: 'Consultants support local staff to develop their competence'. The term 'support' is consciously chosen to cover a very diverse range of activities which are not necessarily easy to compare with each other (such as direct advice, online advisory services, training measures, meetings and conferences, etc.). It might therefore be necessary to use indicators to record single activities and to let these 'speak' for the entire set of activities. For instance, the number of hours of direct advice provided can be easily recorded and its value can represent all areas of advisory services.

Similar considerations apply to the output level: in the 'curricula development' example the output is worded as follows: 'Local staff have the skills to develop curricula'. Again the theoretical construct of 'skills to develop curricula' contains so many elements that it cannot be measured directly. For instance, performance tests might be used during final examinations as an indicator of the local staff's ability to develop curricula: the better the results (marks), the more likely it is that local staff are deemed capable of developing curricula themselves – even if many other competencies are needed for it. There are strong arguments for using this indicator rather than others (e.g. conducting a survey among managers on this topic, which might face resistance from the staff concerned): the necessary data are readily available, they are easy to interpret and they cover a relatively broad range of aspects.

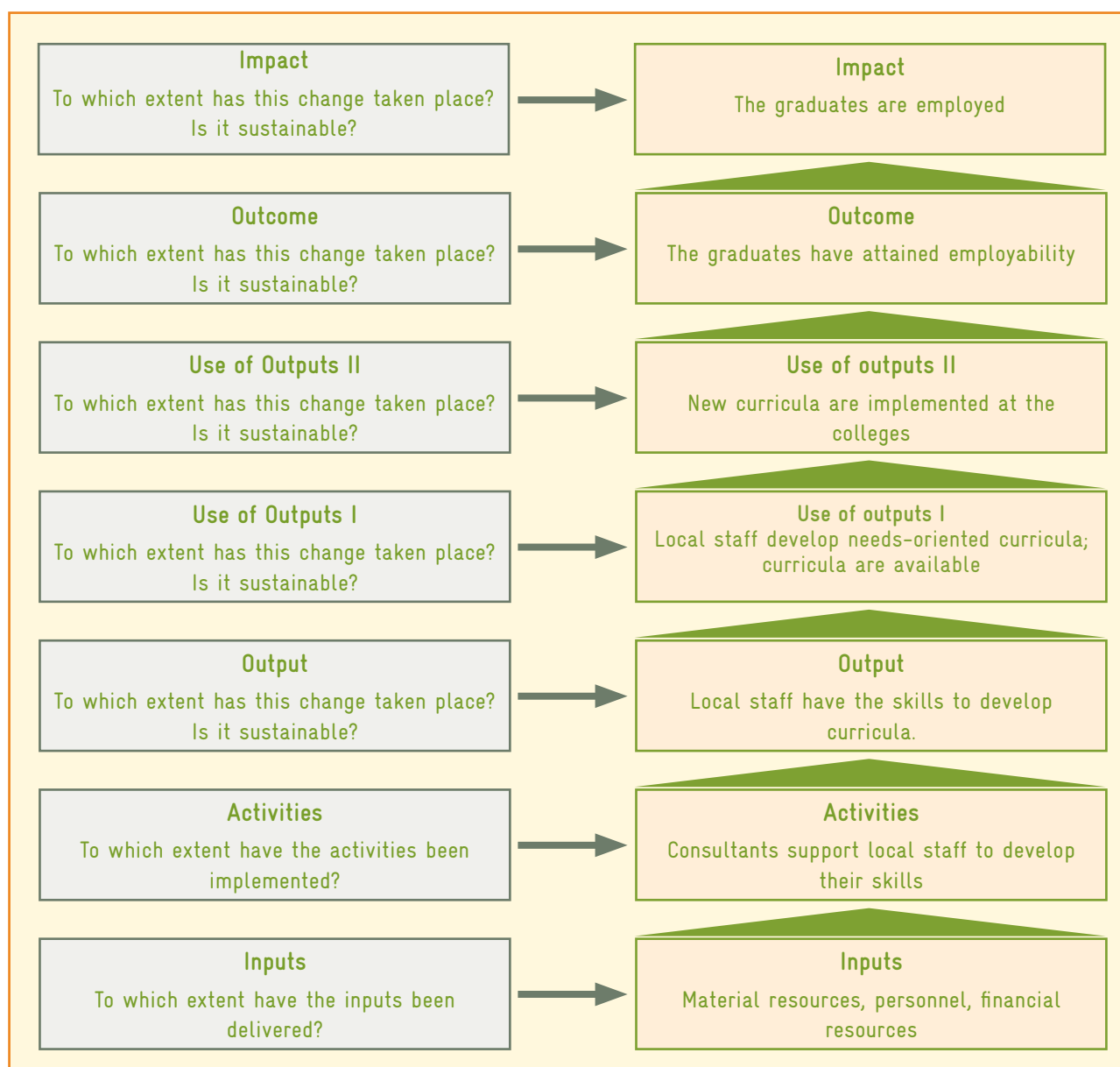
It is particularly important to assess the *use of outputs* that are generated by the project, since it is a precondition for achieving results. Defining good indicators, which adequately reflect this use of the project's outputs, is essential. Our example addresses two levels of use of outputs: the development of new curricula and their implementation. In both cases it may be impossible to collect these data for all the colleges involved with a reasonable amount of effort. The development of vocational training centres

in remote areas, for instance, may only be assessable at a very late stage or not at all; or the total number of curricula may be so large, that it may be impossible to conduct an exhaustive survey; the development of curricula may occur at different places and at different times, etc. One solution may be to select single locations as indicators for the totality of locations, especially if they are well suited to depict the general state of development – which does not necessarily mean the average frequency of use of outputs.

### 3.6 Using indicators for results-based monitoring

Making decisions to steer a project towards success requires *regular information about the issues pertinent to those decisions*. At the moment of deciding whether or not to set a new strategic direction for the project, all this information must be available and be of sufficient quality to serve as the basis for decision making.

**Figure 5: Results chain and core monitoring questions**



A series of indicators that are reliable and suitable for this task need to be selected during the planning stage of the project already. The most important *requirements* are not only the quality of the measuring instruments themselves but above all their *feasibility* (how likely is it that the project itself or its partners will be able and prepared to compile the information needed to apply the indicators?) and their *acceptance* among all parties involved in the decision making process (see chapter 3.4).

Monitoring should thus use indicators that are as simple, understandable, uncontested and reliably available as possible. In the case of results-based monitoring, it is particularly important that it supply conclusions relevant to the results that can be expected from the intervention. This is addressed by developing results hypotheses (see chapter 2.2) as forecasts: in our example the consultants implement a sufficient number of activities to build the competencies for the development of curricula. During the course of the project the activities constitute a necessary precondition and counting them can also be utilised as part of results-based monitoring. Additional circumstances that can foster or hinder the transfer of activities should also be observed during monitoring. These may include legal regulations (for

instance whether the different vocational training centres have the same rights to decide autonomously about the introduction of new curricula) or changes in the labour market (whether there is a demand for the qualifications that are being provided) and should also be measured.

**Figure 6: Monitoring plan for the case study curricula development**

Results level	Objectives	Indicators	Calculation	
Impact	Employment	Proportion of graduates from reformed training programmes who are employed against the total of vocational training graduates	employed graduates/all graduates	
Outcome	Employability	Proportion of graduates from reformed training programmes who have received an offer of employment against the total number of participants in final examinations (for reformed and non-reformed vocational training programmes)	successful graduates with a job offer/all exam candidates	
Use of outputs	Curricula are implemented at vocational training centres	Proportion of vocational training centres, where trainees are trained according to the project's concept against all vocational training centres.	Vocational training centres with training in Y/total amount of vocational training centres in region X	
Outputs	Skills to develop curricula	Proportion of individuals who have passed a performance test on curriculum development against all individuals who have sat for such an exam	tests passed/tests sat for	
...	...		...	

	Numerical value	Source	Instrument; Scope	When & How often
	60 percent of the successful candidates find employment within the first six months after course completion.	Project's own data collection (in writing)	Tracer study; exhaustive survey	Baseline and consecutive yearly updates
	80 percent pass the reformed final test and have received at least one offer of employment within the first two months after course completion.	Project's own data collection (e.g. via email)	Graduates survey; exhaustive survey exams statistics; exhaustive survey	Baseline and consecutive yearly updates
	At least 50 percent of all vocational training centres in the region X that offer training in the area of occupation Y	Regional TVET authority	Public statistics (Reports by vocational training centres); exhaustive survey	Baseline and consecutive yearly updates
	At least 80 percent of all participants pass the certified curriculum developer examination.	Examination statistics (Test participation and test result)	Own statistic; exhaustive survey	Yearly updates
	...	...	...	...

### 3.7 Indicator checklist

Are the indicators in the project proposal defined in such a way that there is no doubt as to what is to be measured?	<input type="checkbox"/>
Have indicators been defined for all steps of the results chain at all levels – Output, Use of output, Outcome, Impact?	<input type="checkbox"/>
Are the proposed indicators SMART?	<input type="checkbox"/>
Are the indicators clearly linked to the actual issue at stake?	<input type="checkbox"/>
Are there indicators to measure sustainability?	<input type="checkbox"/>
Are the indicators at the outcome and impact level aligned with national strategies and indicators to the extent possible?	<input type="checkbox"/>
Does the results chain contain key indicators such as outreach (participants of initial vocational training, participants of further training, education staff, consulting staff, institutions) or employment rate?	<input type="checkbox"/>

## 4. Principles of measurement

Measuring certain aspects of a project is a necessary part of professional project management. The following section deals with the question of requirements concerning the quality of measurements and the criteria to establish when a measurement is 'good enough'. It is ultimately the responsibility of the programme manager to decide how to use the available resources with a view to maximising the usefulness of the measurement for the steering process.<sup>7</sup>

In the context of results-based monitoring, measuring the indicators specified during the planning stage serves as a progress review. If development is not measured and checked regularly and based on the expectations specified at the beginning of the project, deviations or even unfavourable changes are identified too late and cannot be corrected in time. It is not just a question of measuring trends but of also of attribution, in other words of the connection between measured changes and certain causes (usually a certain intervention that was undertaken during the period of observation).

Therefore, the project manager must be aware of when which effects of a given intervention are likely to appear, how these effects can be measured and the degree of precision required in doing so. This is why it is important to define appropriate intermediary steps and objectives in order to be able to gather the information on project progress and results needed for making project steering decisions.

### 4.1 How to assess the quality of measurements?

There are three different yet interconnected requirements on the quality of a measurement: (a) validity, (b) reliability and (c) objectiveness.

The *validity* of a measurement refers to the capacity of a measuring instrument to actually measure what it should measure. A measuring instrument should measure neither more nor less than the object it is supposed to measure, i.e. the deviations from the 'real value' should be as small as possible and independent of other influences. A performance test, for instance, should mirror the performance of a candidate in a certain subject – and not in another subject – and it should provide precisely the result that corresponds to the actual performance – and not that of another candidate.

The *reliability* of a measuring instrument means that using the instrument in the same conditions and the same circumstances always provides the same results. Unlike validity, we are not looking at whether the measurement is 'correct' or whether the indicator was 'applicable' or not, but only at the replicability of the results. A performance test conducted with trainees should always provide the same result if there are no changes to the surrounding conditions or the potential of the candidate.

The third criterion to assess quality of a certain measurement is that of its *objectivity*, i.e. the measurement should be independent of the people conducting it. This, in turn, means that two people should reach the same result when using a certain instrument without consulting each other. This can also be applied to the example of a final test: two teachers or trainers should give the same mark independently of each other.

---

<sup>7</sup> Cf. also the Standard for measuring results in private sector development of the Donor Committee for Enterprise Development (DCED 2010, 2011).

In the particular case of making steering decisions in development projects, the preciseness of a measurement is less important than obtaining clear, timely and accurate information on what needs to be done. Here too demands are often contradictory: a precise measurement usually requires more effort and it takes longer to obtain feedback. Additionally, assessing and interpreting the measurement results is much more difficult and requires specific competencies, which the decision maker does not necessarily have and which may be difficult to transmit to others. Especially when establishing a monitoring system, timeliness, accuracy and intelligibility are therefore more important quality criteria for a measurement (and hence what is required of an indicator system) than achieving a very precise measuring result.

## 4.2 Which particular challenges arise when measuring the results of TVET?

There are a large number of TVET-specific requirements, which some examples will help to illustrate. In most cases it is not a problem of measurement in the actual sense (i.e. difficulties that arise while actually measuring something) but rather lack of clarity regarding what it is that should actually be measured.

For instance, there is no straight forward answer to the question raised in our ongoing example regarding whether a curriculum is ‘needs-oriented’ (use of outputs). Needs assessment implicitly requires *further knowledge* about the (qualifications and training) demanded by the different stakeholder groups (mainly the enterprises, but also government, the TVET institutions in the country, the trade unions and workers associations as well as the employees and trainees). In addition, in the light of economic changes in the highly dynamic economies of transition countries in particular, it should not be assumed that demand will remain stable. Institutional arrangements must be established to ensure that curricula are *continuously adapted* to these changes in order to sustainably achieve high training quality and the desired results in the labour market.

The indicators used to measure these issues cannot possibly mirror their entire complexity and diversity in detail for all companies and situations correctly. This is neither a question of measurement quality nor of the indicators being used; the problem is that the definition of the underlying construct ‘needs-oriented’ is neither precise nor operational enough for measurement.

The ‘*problem of definition*’ becomes even clearer when taking a closer look at the *statements at the outcome and impact level*. In order to be measurable, the statement at outcome level ‘The graduates have attained employability’ needs to be more precise in several aspects:

- ▶ Does ‘graduates’ mean all or would a certain number of them be enough – and how high would that number have to be in order to be considered presentable?
- ▶ Does ‘employability’ refer to the demand for competencies or to the formal recognition of course completion or of the acquired occupational title?
- ▶ Where are the boundaries of the ‘market’ used to ascertain employability? Are we looking at the local labour market, taking a national perspective or does the training have to meet the high demands of the global market? Does ‘market’ refer narrowly to the occupational segment for which qualification was obtained (i.e. a job that matches the type of training) or is it enough to increase chances of employment regardless of the work actually being done?

Requirements can also change rapidly, while conducting training and adapting curricula requires a certain amount of time. This is why at the outcome level there is always a certain '*problem of attribution*' (in the sense of the responsibility of the project or programme for quickly adapting to changes in demand), which obviously becomes more noticeable at the impact level.

The economic results mentioned in the example ('graduates find employment') depend on a number of additional factors, such as the general economic development of the country, the employment practice of enterprises or the individual behaviour of the employees over which the project or programme has little to no influence. These are questions of *attribution*, to which metrological answers exist (see chapter 5).

Then there is the question of the scope of responsibility, in other words for how long the project is to assume responsibility for the career path and economic development of a worker's household. Does implementing a needs-oriented curriculum automatically mean that the project assumes long term responsibility for the occupational and economic prospects of its trainees or are there limits that can be clearly defined? This question regarding the (potential) *sustainability of results* points to the hypothetical sequence of results, for instance the process of 'forgetting' what has been learned after completion of training. In order to be able to compare the actual development with what was expected – and thus to assess the durability of an effect achieved during the course of the project – it is necessary to develop assumptions regarding how this process would have taken place without further interventions or disruptions.

These and similar questions need to be answered in a TVET intervention if an adequate results-based monitoring system is to be implemented. The more accurately the objectives are defined, the more accurately can the aspects to be measured be narrowed down and assessed by means of appropriate indicators. The effort of measurement does not necessarily have to be large: **simple and creative solutions that offer sufficient information for decision making are better than complicated and time-consuming measurement procedures with an unnecessarily high level of precision.** This is also true of the assessment of causality assumptions, which plays a decisive role in attribution in the results measuring context and helps to indicate whether the measures conducted have achieved the desired results or not.

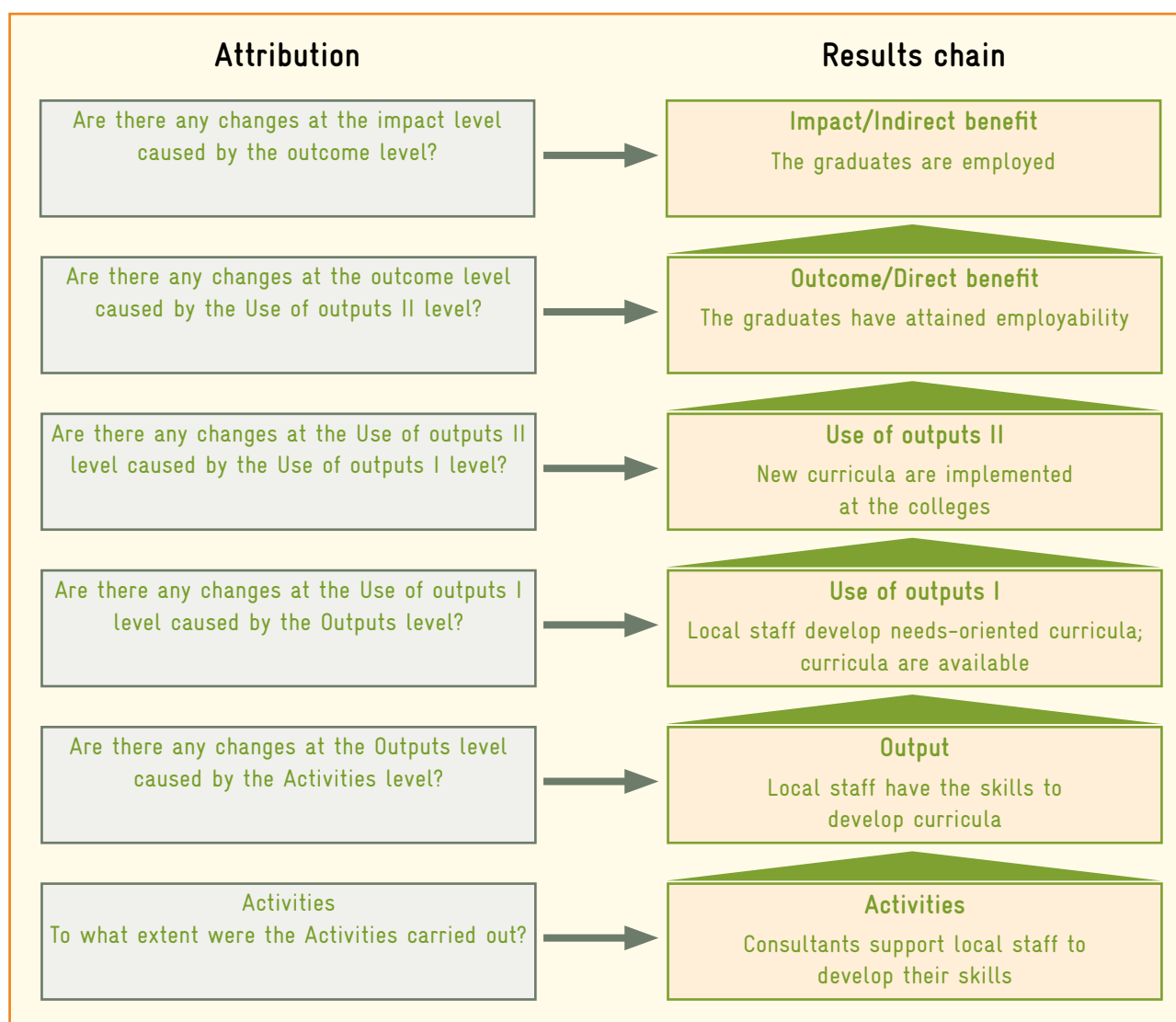
### 4.3 Measurement of indicators checklist

Does the project or programme have a plan containing the following information for all the indicators:	
▶ Which data need to be collected to measure the indicator?	<input type="checkbox"/>
▶ Which measurement procedures are to be employed?	<input type="checkbox"/>
▶ At which time intervals are measurements to take place?	<input type="checkbox"/>
▶ Who is responsible for implementing data collection and who is responsible for ensuring the quality of data?	<input type="checkbox"/>
Are the measurement procedures cost-efficient? Are there better alternatives, which might provide at least a similar quality of data?	<input type="checkbox"/>

## 5. Measuring interrelation of results

Even if changes occur and can be measured, it does not necessarily mean that these changes were in fact sparked by the project in question. In the case of our example, it might well be that all the graduates of the reformed training programmes find employment. This might be due to a significant improvement in the quality of training as a consequence of the introduction of new curricula. But the positive employment effect might also be due to a general economic recovery. Changes are often caused by several factors.

**Figure 7: Results chain and attribution of changes**



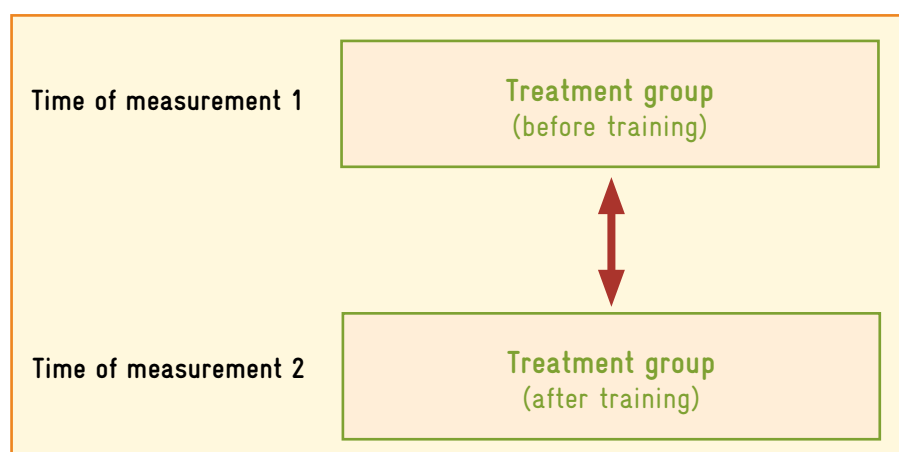
Attributing a change to one (or several) cause(s) means we regard the change as being caused by it (them). Attribution is always based on a cause-effect-assumption (causal assumption), which needs to be verified. It follows the results hypotheses. Attribution affects all elements of the results chain and there are several ways of verifying attributions.

If the requirements described in the previous chapters – results chains, indicators, baseline, and correct measurement of indicators – are fulfilled, then we can use attribution to measure the interrelation of results.

## 5.1 Methods to verify the interrelation of results

Usually every development project assumes that the measures undertaken will bring positive change for the target groups. If, for instance, a new training concept is introduced during the course of an intervention, this is done to achieve improvements to the competence of the participants. The simplest way of verifying this assumption is to measure the competence before and after the intervention and to compare these two results. The measurement taken before the intervention reflects competence without intervention, i.e. the counterfactual. The ‘counterfactual’ is generally understood as the situation which would have prevailed if no action had been taken. Hence, it deals with the question of ‘what would have been if’. In our example we assume that the initial condition would have remained stable during the course of time if no intervention had taken place.

**Figure 8: Simple before/after comparison without control group/comparison group**



Such a simple before/after comparison has the disadvantage that the changes measured might have been influenced by other factors – for instance, the participants may have attended other courses. Because there is no control group, the simple before/after comparison cannot prove the causal relationship between the changes observed and the activities carried out by the project. This procedure can still provide plausible evidence on the project’s results, for instance in the case of conducting short courses or vocational further training (the participants can be surveyed, for instance, regarding their employment situation before and after the training – possibly also retrospectively to a certain point in time).

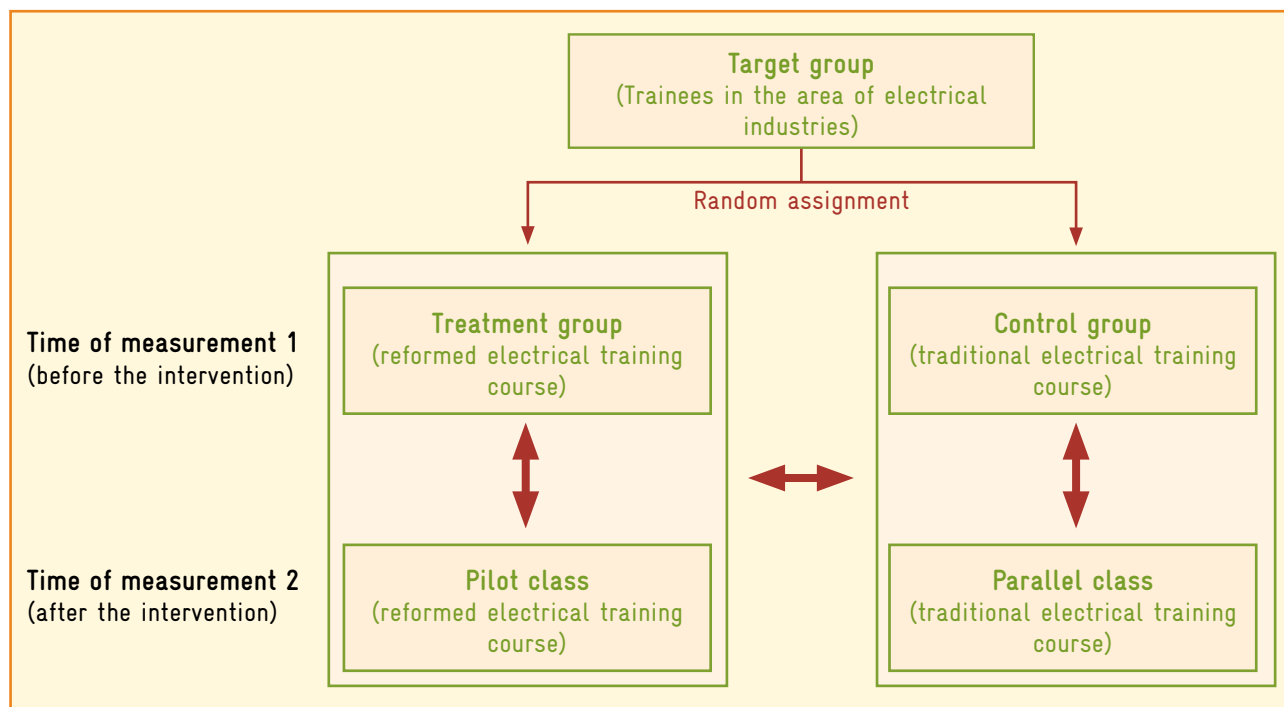
In the following sections we present four classic strategies for verifying the interrelation of results; (a) Experimental design; (b) Quasi-experimental design; (c) Cross-sectional comparisons and (d) Qualitative design. Chapter 5.2 shows an application of all four verifying strategies to the example ‘curriculum development’.

### 5.1.1 Experimental design

The core idea of an experiment is to randomly distribute the inevitable survey errors (such as differences in the participants' motivation) across two groups and to subject one of the groups only to an intervention. If the intervention takes effect, then the results of the two groups must differ to a statistically significant degree. Provided the overall circumstances for both groups are equal and the individual factors of the participants are mathematically controlled by random assignment, the difference between the groups can be interpreted as a causal effect of the intervention.

Therefore, an experimental design is a very good way to verify whether we can infer that a certain intervention has caused a certain change. The procedure is easy to describe in the context of education measures: for instance, if a new education concept is to be tested, we can establish two groups. While one group is trained following the concept used so far, the training of the other group follows the new concept. The participants are randomly selected and assigned to one of the groups (randomisation) prior to the intervention – in other words the implementation of the training. Before and after the training a performance test is conducted, which can provide information about changes the intervention could bring about. Because of the random assignment the group results can now be mathematically separated from the effects of the intervention and conclusions drawn on the causality.

**Figure 9: Example for an experiment**



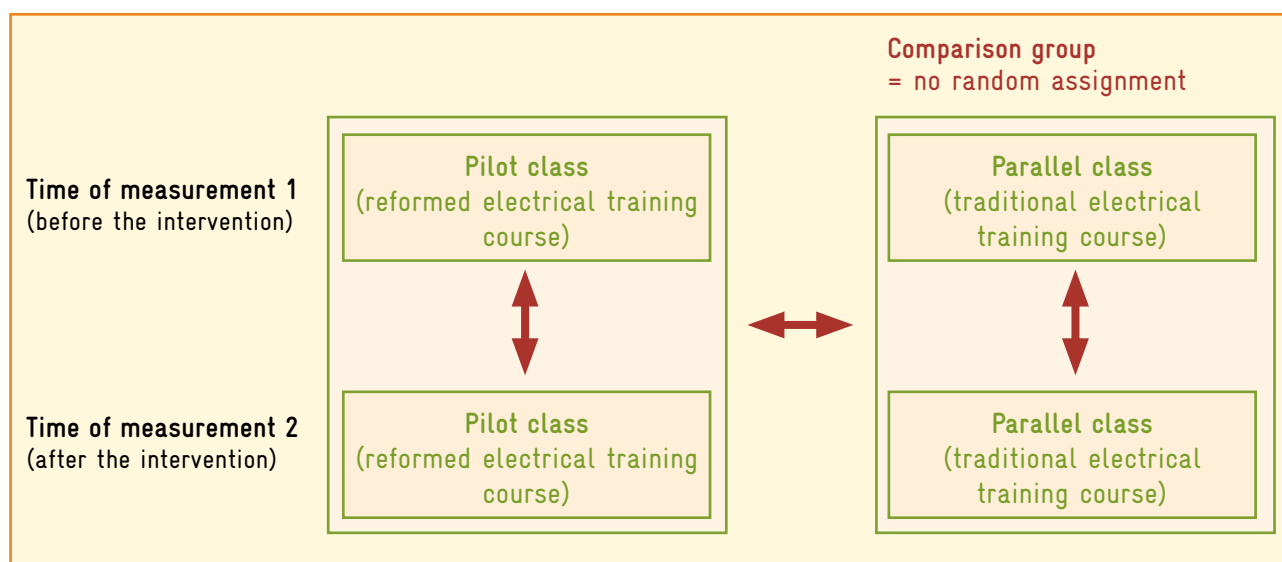
It is important to note at this point that for such an experimental design the randomisation is more relevant to the results than the number of participants. Usually group sizes of about 20–30 people per course are quite sufficient for an experiment and for checking the interconnection of results it sets out to verify.

### 5.1.2 Quasi-experimental design

In practice it is often difficult or even impossible to establish two groups by random assignment and to keep the conditions of both groups constant between the before and after measurements. If, for instance, two training classes are established for an experiment, not only the participants must be randomly assigned to the group but all other factors (e.g. trainers, rooms, teaching material, etc.) must be the same or as similar as possible, so that the differences measured between the two groups and the times of measurement can be conclusively attributed to the intervention (attribution). In the case of our example, this means that the framework conditions for both groups would have to be controlled for the duration of a whole training period – an ambitious task, not just in developing countries. It can be generally said that the probability of not being able to fulfil the optimal conditions for an experiment increases with the complexity of the influencing factors and the time span between the measurements.

In this case it is common to speak of a ‘quasi-experiment’, which differs from an experiment in the limited control and the non-random assignment of the participants in both experimental groups. The two groups to be compared are selected following theoretical considerations with the aim of having two groups that are as similar as possible to each other and only differ in terms of the intervention. As opposed to an experiment, these are real, existing groups, which are not artificially established for the sole purpose of the assessment. This may indeed also prove an advantage compared to the experiment: the members of the groups are not aware of the fact that they are part of an experiment and therefore will not try to manipulate it. Therefore the ‘quasi-experiment’ should not be regarded as just an ‘incomplete’ experiment but as a procedure in its own right with its own specific advantages (being closer to reality and less prone to being manipulated) and disadvantages (being more difficult to control and more likely to lead to attribution errors).

**Figure 10: Example for a quasi-experiment**



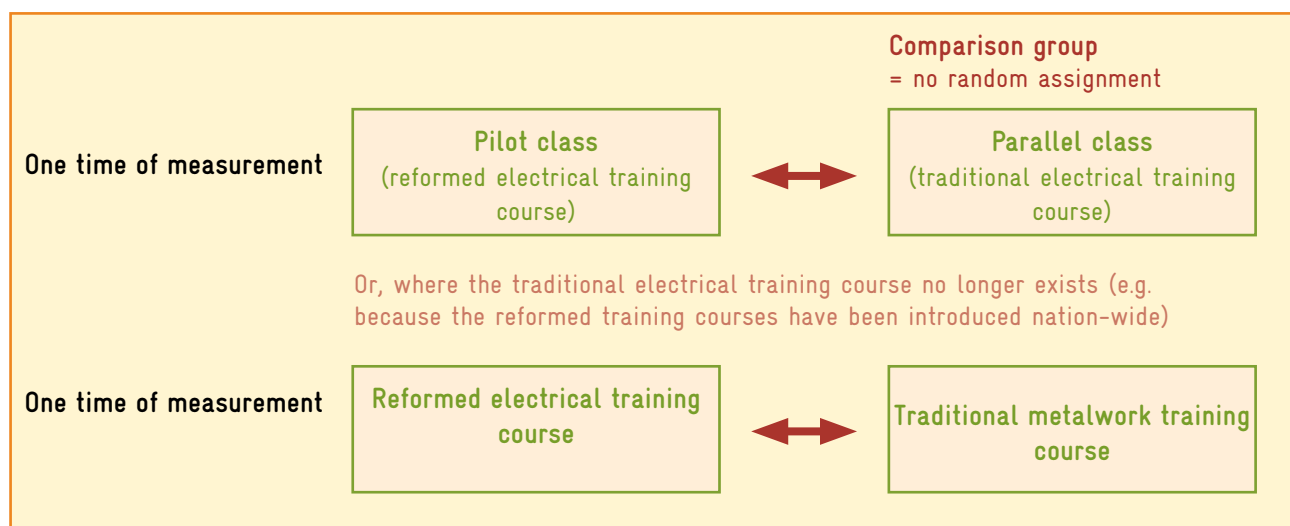
By relating the results with a comparison group (e.g. a parallel class) it is possible to verify the causality in a similar way as it would be done in an experiment. The lack of a random assignment before the intervention, however, means that the deviation in the results of the different groups cannot be removed by statistical procedures. For instance, systematic group bias may occur because of the procedure used

to select the groups: if only the best students are selected to be part of the class which will be using the new training methods, we must assume that this assignment will influence the performance tests before and after the intervention. One possible result might be that the learning progress is less than in the comparison group because the level of competence was already higher before the intervention and the potential scope for progress from the new teaching methods correspondingly lower. If such influences are known – or as shown in the example are already visible at the beginning – they can be controlled on the basis of theoretic assumptions. For instance, the assumption that better students achieve a smaller increase in competences could be verified by comparing them with a third group who had a level of competence above the average prior to the intervention. This shows how non-adherence to the requirements of an experimental design means more effort to control interfering factors. In practice, we need to assess which effort – ensuring experimental design conditions or controlling possible interfering factors – is smaller and therefore easier to handle.

### 5.1.3 Cross-sectional comparison

In many aspects cross-sectional comparison is similar to quasi-experimental design. In this kind of survey, the ‘non-participants’ act as a comparison group, which mirrors the counterfactual condition of the participants. For instance, all students of a particular vocational training centre (or a random sample) may be surveyed using a questionnaire and then the participants of a reformed training course could be compared with the other students. However, even in the case of a random sample or an exhaustive survey it cannot be assumed that students were assigned randomly to the treatment or comparison group, since this decision was made before the survey either by the participants themselves or by others – just as in the case of a quasi-experiment.

**Figure 11: Example for a cross-sectional comparison**



Compared to the quasi-experimental design, the cross-sectional comparison has the disadvantage that there is no before-measurement (baseline). Therefore it will not be possible to clearly ascertain whether there has been any progress at all: if differences are identified between the participants and the non-participants, these may have existed prior to the training. A lack of baseline data always leads to uncertainty whether progress has been achieved at all. Matching procedures can be used to simulate the counterfactual condition by establishing statistical twins (i.e. the comparison between two cases that are largely

equal with regard to as many attributes as possible, such as age, sex, regional and social origin, educational level, parents' income, religious affiliation, tribal origin, etc. and only differ with respect to the participation in the reformed training course). This cannot replace the baseline data, but in some cases it can provide good information regarding a potential causality.

Matching procedures are well suited to analyse changes affecting relatively stable aspects that are known to be strongly influenced by the selected features. It has been proved by many surveys, for instance, that limited education, poor language proficiency, health risks, and gender (as well as a number of other features) have a strong influence on people's employment prospects. In this case it makes a lot of sense to establish a 'statistical twin' who has similar characteristics but has not participated in the intervention. Matching procedures are less suitable, though, for measuring attitudes, which usually are only very slightly linked to person-specific variables such as age, sex and education.

#### 5.1.4 Qualitative design

Qualitative procedures are particularly suited to measure objects that are difficult to narrow down. In this case, qualitative procedures can help understand how certain aspects relate to each other. In other words and as opposed to the previously described methods, they are less suited to verify causality hypotheses than to infer the meaning of certain causal connections.

Take the following example: a project's aim is to reintegrate former child soldiers into society by offering them vocational training. The vocational training is therefore not just a means to improve the participants' chances on the labour market in order to obtain secure employment, but to a much greater extent it provides a sense of purpose and an opportunity to 'heal' the dysfunctional relationship between the individual and society. The constructs mentioned – 'social integration' and 'sense of purpose through work' – are not clearly appreciable and are shaped by the individual perspective of the affected child soldiers and their socio-cultural circumstances. While the question regarding whether someone does or does not find employment that adequately matches their vocational training can be largely answered 'objectively', this is not true of the categories mentioned above as they involve things like the personal feelings of those involved. Such aspects can only be explored from the individual perspective and the causal mechanisms between intervention and results (in other words the relevance of vocational training for developing a 'sense of belonging') can at best be comprehended and understood on a case by case basis.

In these cases qualitative procedures can prove useful because they are methodologically open. This may be done through individual interviews, which can help to put oneself in the place of the persons affected and to understand their emotions and behaviour, but also in group discussions, where results are developed jointly and questions are discussed, for instance regarding the respondents' relationship with society. In these examples the discussions are not just geared towards gathering data but they may also serve a therapeutic purpose. Of course, qualitative procedures may also be used without this connection, for instance in the form of participant observation, which can provide information regarding the social behaviour of trainees.

The main difference between qualitative and quantitative designs is that the first are much closer to everyday life. For development cooperation, though, they pose a special methodological challenge: it is crucial to know the socio-cultural circumstances and to take them into account when assessing the procedures and the results they provide. The use of certain procedures such as focus group discussions may

be highly problematic in certain cultural contexts, while in others it may be a really valid and useful instrument. In the previous example it may even have psychological side effects and put the young people under additional stress. Thus, ‘understanding’ a country’s culture is not only crucially important for the implementation of the interventions but also for selecting the survey and interpretation procedures. Applying the analyst’s own cultural understanding can lead to false interpretations, especially in the case of qualitative methods.

## 5.2 Verifying the interrelation of results using curriculum development as an example

The following tables present different attribution strategies and methods for collecting data for the four higher level of the results chain for curriculum development. This description does not necessarily cover all aspects. The strategies and methods mentioned are not only applicable to the example of curriculum development; they can actually be used in different TVET-related contexts (see also chapter 5.3). In practice, the decision on which survey strategy to use at which level will ultimately depend on the personnel and financial resources available in the project and the local conditions and opportunities.

Level in the results chain	Recommended methods for data collection
<p><b>Impact</b></p>  <p>Usually the issues at stake are:</p> <p><b>Employment</b><sup>8</sup> in terms of gainful employment and decent work as well as equitable participation in social and economic life.</p> <p><i>It must be considered at this level that the graduates may consciously choose to proceed to further education, e.g. higher education. This can have social advantages and be worthwhile economically. Whether these transitions are valued as positive, neutral or negative depends on the exact objectives and the chosen indicators. In case of doubt, these target groups should also be considered for the purposes of results-based monitoring.</i></p>	<p><b>At the participants' level:</b></p> <p>Tracer Study covering all or a selection of graduates. This may be used to trace other economic and social impacts beyond the actual employment. Weak point: very expensive in countries with poor infrastructure. Tracer studies are too costly for short courses.</p> <p><b>Survey</b> among the participants about the benefit/advantages that the acquired competencies have when seeking employment. Weak point: subjective perception leads to distortions.</p> <p><b>At labour market level:</b></p> <p><b>Survey</b> among staff of employment agencies about the participants' placement ratio in the labour market in comparison with other job seekers.</p> <p><b>Survey among entrepreneurs or managers who employ graduates or interns of reformed training courses.</b></p> <p><b>Survey</b> among customers/clients of self-employed graduates, in order to compare their performance with that of their competitors.</p> <p>Weak point in all cases: None of the groups has comprehensive knowledge (Bias due to subjective perspective).</p> <p><b>At the level of the employment system:</b></p> <p><b>Statistical data</b> on employment and unemployment. Weak point: official data are often not differentiated enough, issued too late and not reliable, especially with regard to hidden or partial unemployment.</p>

<sup>8</sup> With regard to measurement of employment effects, cf. 'Measuring employment effects of technical cooperation interventions – Some methodological guidelines' (Kluve et al. 2010).

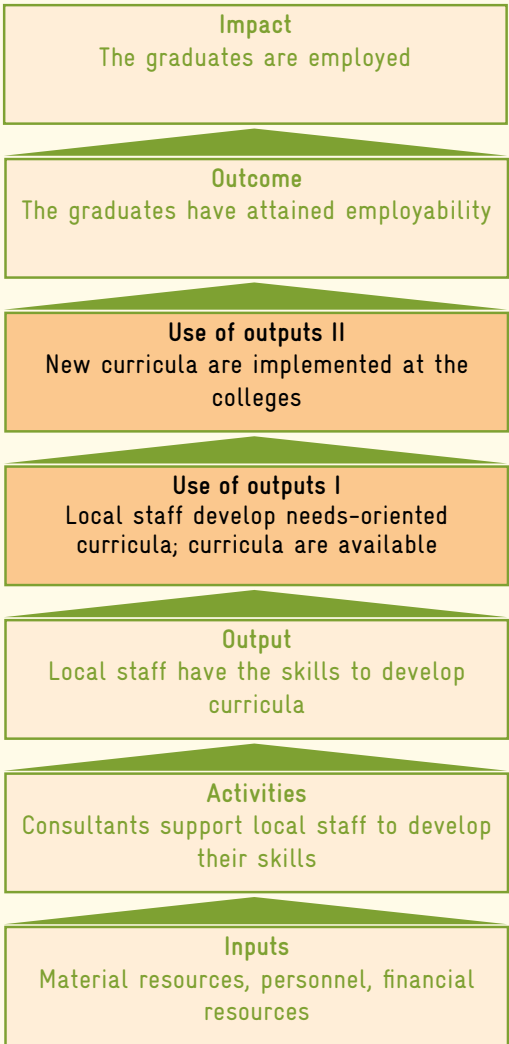
## Possible methods for verifying causality ('attribution')

- A. **Experimental design** Comparison of an treatment group with a control group (of 30 people each). Both groups are established based on a random procedure (Randomised Controlled Trial, RCT). Another aspect of the experimental design is the comparative measurement at two different points in time: the baseline and after the intervention. If the participants of a reformed training programme are selected randomly before the programme starts (and not according to performance criteria, for instance), it is possible to compare with a parallel class being trained following the previous concept. What is important is that both groups are then treated exactly the same, which allows the employment effect to be verified (→ see also 'experimental design')
- B. **Quasi-experimental design** This requires at least one comparison group, for instance graduates who are trained simultaneously following the previous curriculum. As opposed to the control group, the participants of comparison group are not assigned randomly.  
If data on the employment records of several cohorts of graduates who were trained following the old curriculum in previous years are compared with the data of graduates of the reformed training courses, then it is actually a longitudinal study. (→see 'quasi-experimental design').

Level in the results chain	Recommended methods for data collection
<p><b>Outcome</b></p> <p><b>Impact</b> The graduates are employed</p> <p><b>Outcome</b> The graduates have attained employability</p> <p><b>Use of outputs II</b> New curricula are implemented at the colleges</p> <p><b>Use of outputs I</b> Local staff develop needs-oriented curricula; curricula are available</p> <p><b>Output</b> Local staff have the skills to develop curricula</p> <p><b>Activities</b> Consultants support local staff to develop their skills</p> <p><b>Inputs</b> Material resources, personnel, financial resources</p> <p>Usually: <b>Employability</b> in terms of comprehensive professional capability to act.</p>	<p><b>At the participants' level:</b></p> <p><b>Final or intermediary exams</b> or competence assessments in terms of measuring the acquired competencies as compared to the learning objectives specified in the curriculum. This is based on the assumption that the learning objectives are relevant for employment, which was ensured at the lower levels of the results chain. The participants' successful demonstration of the relevant competencies can be equated with the construct of employability. Weak point: This is based on the assumption that employability is the only criterion used to award marks and that other aspects (such as teaching and learning skills, employers' preferences) do not cause systematic distortions of the results.</p> <p><b>Survey</b> among the participants about the benefit/advantages that the acquired competencies have when seeking employment or pursuing further training, e.g. higher education. Weak point: subjective perception.</p> <p><b>At labour market level:</b></p> <p><b>Survey</b> among staff of employment agencies about the participants' placement ratio in the labour market in comparison with other job seekers. Weak point: Bias. Not all graduates use these services.</p> <p><b>Survey</b> among training staff in enterprises and training institutions in order to estimate the employability of trainees and/or interns belonging to reformed training courses. Weak point: Snapshot, which may change over time.</p>

## Possible methods for verifying causality ('attribution')

- ▶ **Cross-sectional comparison** Comparative analyses can also be conducted without measuring at two different points in time. For instance, one measurement taken after the intervention can be used to compare the employability of people who have been trained following the reformed concepts and those trained following the traditional concepts, even if no baseline study was conducted. Another possible application results when a certain training area (e.g. electrical training) was reformed in the entire region covered by the study and there are no graduates of traditional training course available as a comparison group. Graduates of traditional training courses in a different but related training area (e.g. metalwork at a comparable educational level) might be used as a comparison group. Compared to the experimental design, the cross-sectional comparison is similar to the quasi-experimental design in that the assignment of the comparison group is not random but follows certain theoretical considerations. 'Statistical twins' are established with the assistance of matching procedures (e.g. Propensity Score Matching), so that the effect of the intervention can be estimated mathematically and thus causality can be verified to a certain extent.  
(→ see '**Cross-sectional comparison**')
- ▶ **Longitudinal studies** with their repeated measurements make it possible to estimate the development trends for employability. Usually they compare the situation before and after the intervention. If the measurements after the intervention provide consistently better values than before the intervention, then this allows inferring the existence of a causal relationship between this developmental change and the intervention. This may be done using aggregated data (trend survey) as well as individual data (panel survey).  
**Trend surveys** for instance can use average marks in final or intermediary exams to ascertain to which extent the employability of participants of reformed training programmes has improved collectively. It is important to ensure that the tests are adapted on a regular basis, since employability is not an absolute and constant trait. It evolves as the employers' requirements change.  
**Panel surveys** can be used to follow individual education and employment paths and provide information about successes in acquiring competencies and the process of attaining employability. This may be used to ascertain who are helped by the new didactic approaches set out by a reformed training course and who not.

Level in the results chain	Recommended methods for data collection
<p><b>Use of outputs</b></p>  <p>Usually: Qualitatively and/or quantitatively <b>improved Vocational training programmes</b>, as well as an improved access to these programmes for a wider range of people.</p>	<p><b>At the level of the training staff:</b></p> <p>The application of new curricula aims at improving the quality of teaching and learning. In this case, the first thing to look at is whether the newly developed curricula are actually applied or not-assuming that the curricula fulfil the quality standards in terms of content and form. The main aspect here is to understand the reasons stated by the people involved to justify why the new curricula are or are not being applied. Therefore, it is necessary to gather their subjective opinions and the logic of their argumentation. This can be done in different ways, e.g. by conducting expert interviews with teachers and trainers on the degree of obligation to apply the curricula, their self-assessment regarding their technical and methodological competence as well as their motivation to implement the curriculum, their assessment of the learning environment and the materials available, the attitude of the school and enterprise management as well as the trainers' and teachers' view on the consequences for this arising from the application of the curricula.</p> <p>Weak point: to get a comprehensive view, it may be necessary to conduct and carefully evaluate a large number of long and intensive discussions.</p> <p>Additionally <b>Auditing/Observation</b> can be used to directly observe the application process at the learning sites and be in a better position to appraise it. Weak point: participant observation can influence the behaviour of people and distort the results.</p> <p><b>Focus group discussions with school directors or heads of department of the affected vocational schools or training programmes might also be suitable. Weak point: the group dynamics can lead to different and distorted results.</b></p>

## Possible methods for verifying causality ('attribution')

- A. **Qualitative design** In comparison with the previously described procedures, the qualitative procedure is geared to a larger extent towards single cases and tries to explore their specific features. Hence, the selection of cases, for instance, follows theoretical considerations and not statistical principles and the analysis tries to explore as comprehensively as possible the characteristic traits of a certain case instead of trying to verify pre-established hypotheses. This is particularly useful when assessing the introduction and testing of innovations (new didactic concepts, new technologies, new advisory methods, etc.) – and therefore especially suited for the use-of-outputs level. During the testing phase of a new training programme it is particularly interesting to understand the reasons why this programme is being accepted or rejected, rather than a quantitative indication of the degree of acceptance or rejection. The point is to gather and understand the lines of argumentation for and against the vocational training programme as completely and with the greatest detail as possible.  
(→see 'Qualitative design').

## 5.3 Further examples for the application of strategies and data collection methods to verify the interrelation of results

### 5.3.1 Optimizing the organisational and management structure of vocational schools

Level in the results chain	Methods for verifying causality ('attribution')	
<b>Impact</b> (Employment)	<ul style="list-style-type: none"> <li>▶ (Quasi)-experimental design</li> <li>▶ Cross-sectional comparison</li> </ul>	
<b>Outcome</b> (Employability)	<ul style="list-style-type: none"> <li>▶ Market-Player-Opinion</li> <li>▶ Qualitative design</li> </ul>	
<b>Use of outputs</b> (Use of the school directors' new competencies: introduction of new organisational and management concepts in vocational schools)	<ul style="list-style-type: none"> <li>▶ Qualitative design</li> </ul>	
<b>Output</b> (Increased competence of school directors with regard to school organisation and management)	<ul style="list-style-type: none"> <li>▶ Qualitative design</li> </ul>	
<b>Activities</b> (Further training of school directors on school organisation and management)	<ul style="list-style-type: none"> <li>▶ Trend survey</li> </ul>	

	Recommended methods for data collection	Key questions
	<ul style="list-style-type: none"> <li>▶ Survey among students of vocational schools where the new organisational and management concepts were applied/not applied.</li> </ul>	<ul style="list-style-type: none"> <li>▶ Is the employment situation for graduates of reformed vocational schools, for instance in a pilot district, better than for graduates of schools that have not been reformed?</li> </ul>
	<ul style="list-style-type: none"> <li>▶ Guided interviews with selected students of vocational schools where the new organisational and management concepts are being implemented.</li> <li>▶ Focus group discussions with entrepreneurs who employ interns/graduates from reformed vocational schools.</li> </ul>	<ul style="list-style-type: none"> <li>▶ Has the implementation of the new organisational and management concepts led to an improved training quality and therefore enhanced employability?</li> </ul>
	<ul style="list-style-type: none"> <li>▶ Participant observation, records</li> <li>▶ Guided interviews with school inspectorates</li> </ul>	<ul style="list-style-type: none"> <li>▶ Are the new organisational and management concepts being implemented by the school directors at their schools? Do school directors who have not participated in the further training also introduce new organisational and management concepts? Is the school directors' initiative to introduce new organisational and management concepts a response to their increased competence/motivation?</li> </ul>
	<ul style="list-style-type: none"> <li>▶ Guided interviews with school directors</li> </ul>	<ul style="list-style-type: none"> <li>▶ Has competence increased? Is increased competence a result of the further training imparted by GIZ/consultants?</li> </ul>
	<ul style="list-style-type: none"> <li>▶ Statistical data on participants</li> </ul>	<ul style="list-style-type: none"> <li>▶ Have enough people participated in the further training? How many schools did they come from? Are the areas of school organisation and management sufficiently covered in these schools?</li> </ul>

### 5.3.2 Optimizing career guidance, employment services and training in employment agencies

Level in the results chain	Methods for verifying causality ('attribution')	
<b>Impact</b> (Employment)	<ul style="list-style-type: none"> <li>▶ (Quasi)-experimental design</li> <li>▶ Cross-sectional comparison</li> </ul>	
<b>Outcome</b> (Employability)	<ul style="list-style-type: none"> <li>▶ (Quasi)-experimental design</li> <li>▶ Cross-sectional comparison</li> <li>▶ Market-Player-Opinion</li> </ul>	
<b>Use of outputs</b> (Use of new competencies by staff of employment agencies)	<ul style="list-style-type: none"> <li>▶ (Quasi)-experimental design</li> <li>▶ Cross-sectional comparison</li> </ul>	
<b>Output</b> (Increased competence of staff of employment agencies in district X)	<ul style="list-style-type: none"> <li>▶ Experimental design</li> </ul>	
<b>Activities</b> (Further training of staff of employment agencies in district X)	<ul style="list-style-type: none"> <li>▶ Trend survey</li> </ul>	

	Recommended methods for data collection	Key questions
	<ul style="list-style-type: none"> <li>► Evaluation of official statistics on employment development from those employment agencies' which have/have not introduced the new concepts</li> </ul>	<ul style="list-style-type: none"> <li>► Is the employment situation for students of lower and higher education, youth, unemployed persons, interns in a pilot district better as for those in a comparison district?</li> </ul>
	<ul style="list-style-type: none"> <li>► Representative survey of students of lower and higher education, youth, unemployed persons, interns in selected districts where the employment agencies have/have not introduced the new concepts</li> <li>► Guided interviews with agents and advisers from selected agencies</li> </ul>	<ul style="list-style-type: none"> <li>► Has the employability or the capacity of youth and unemployed persons to orient themselves (e.g. in the labour market) improved due to the implementation of the new counselling, placement and training approaches?</li> </ul>
	<ul style="list-style-type: none"> <li>► Written representative survey of users of advisory services in employment agencies which have/have not introduced the new concepts</li> <li>► Interviews with directors of employment agencies</li> </ul>	<ul style="list-style-type: none"> <li>► Are new approaches to counselling, placement and training implemented in the employment agencies in district X? Do employment agencies in other districts also introduce these new concepts?</li> <li>► Is the introduction of new approaches to counselling, placement and training a response to the increased competence/motivation of the employment agencies' staff?</li> </ul>
	<ul style="list-style-type: none"> <li>► Test sessions in advisory services before and after the further training imparted by the consultants (before-after-measurement)</li> <li>► Comparative test sessions in advisory services provided by advisers who have not participated in the further training (only measured after the intervention)</li> </ul>	<ul style="list-style-type: none"> <li>► Has the competence increased? Is the increased competence a result of the further training imparted by GIZ/consultants?</li> </ul>
	<ul style="list-style-type: none"> <li>► Statistical exam records</li> </ul>	<ul style="list-style-type: none"> <li>► Have enough advisers from the employment agencies in district X passed the test? Have they acquired the desired competencies?</li> </ul>

### 5.3.3 Optimizing the entire national TVET system

Level in the results chain	Methods for verifying causality ('attribution')	
<b>Impact</b> (Employment)	► Cross-sectional comparison	
<b>Outcome</b> (Employability)	► (Quasi)-experimental design ► Qualitative design	
<b>Use of outputs</b> (Introduction of reformed TVET programmes)	► Qualitative design	
<b>Output</b> (Understanding of a joint responsibility for TVET)	► Qualitative design	
<b>Activities</b> (Advisory services to government departments, enterprises, interest groups, etc.)	► Qualitative design	

	<b>Recommended methods for data collection</b>	<b>Key questions</b>
	<ul style="list-style-type: none"> <li>► Evaluation of official statistics on employment development of reformed/non-reformed occupations</li> </ul>	<ul style="list-style-type: none"> <li>► Is the employment situation for graduates of reformed TVET programmes better than for graduates of non-reformed programmes?</li> </ul>
	<ul style="list-style-type: none"> <li>► Expert interviews with entrepreneurs who employ people trained in the reformed/non-reformed system. Expert interviews with entrepreneurs who do/do not train participants themselves.</li> <li>► Expert interviews with directors of schools involved in the reform process</li> <li>► Expert interviews with independent TVET experts</li> </ul>	<ul style="list-style-type: none"> <li>► Has the quality of TVET improved due to the implementation of the reformed TVET system and has this improved the graduates' employability?</li> </ul>
	<ul style="list-style-type: none"> <li>► Network analysis</li> <li>► Participant observation, (participation in meetings)</li> <li>► Expert interviews with TVET experts</li> <li>► Evaluation of documents</li> </ul>	<ul style="list-style-type: none"> <li>► Have committees been created for the exchange of expert knowledge? Do all relevant actors participate regularly? Do they discuss important topics? Do they implement joint decisions? Do these lead to a nationwide reformed TVET? Where does practical training take place?</li> </ul>
	<ul style="list-style-type: none"> <li>► Expert interviews with the participants of advisory services</li> <li>► Expert interviews with managers of the same organisation who have not received advisory services</li> </ul>	<ul style="list-style-type: none"> <li>► To what extent has the stakeholders' understanding of TVET changed? Is this understanding disseminated within the organisations? Have the recommendations been implemented? Can this development be traced back to the advisory services provided by GIZ/the consultants?</li> </ul>
	<ul style="list-style-type: none"> <li>► Lists of participants</li> <li>► Guided discussions with the advisors</li> </ul>	<ul style="list-style-type: none"> <li>► Have the right people received advisory services? Were all stakeholder groups sufficiently represented? Were the recommendations accepted?</li> </ul>

## 6. Final remarks: 10 tips for successful results-based monitoring

The following comments highlight again the most important aspects of successful results-based monitoring:

- ▶ Projects must be conceived based on realistic results hypotheses.
- ▶ Appropriate indicators need to be used at all levels of the results chain (process and result indicators).
- ▶ A control or comparison group should be identified at the beginning of the project.
- ▶ Baseline studies should be conducted for the treatment and the comparison groups.
- ▶ A results-based monitoring system should be established which makes it possible to gather data on a continuous basis and to do comparative analyses.
- ▶ Secondary data should be identified, checked and used (national statistics, etc.).
- ▶ Appropriate attribution strategies and data gathering methods should be used.
- ▶ The results of results-based monitoring should be used for shared learning – within the project itself but also beyond it with view to future interventions.
- ▶ Results-based monitoring should be carried out jointly with national and international organisations as well as other donors.
- ▶ The local capacity for monitoring and evaluation should be strengthened in the partner country.

## 7. Sources of reference and further reading

- BMBF** (2005): Die Reform der beruflichen Bildung. Berufsbildungsgesetz 2005 (Artikel 1 des Gesetzes zur Reform der beruflichen Bildung – BerBiRefG). Zusammenstellung der Begründungen zu den Einzelschriften des Berufsbildungsgesetzes (Auszug). BMBF. URL (13.11.2009): [http://www.bmbf.de/pubRD/begruendung\\_BBiG\\_einzelschriften.pdf](http://www.bmbf.de/pubRD/begruendung_BBiG_einzelschriften.pdf)
- BMZ** (2005a): Technical and Vocational Education and Training and the Labour Market in Development Cooperation. Strategies 140. Bonn: BMZ.
- BMZ** (2005b): Berufsbildung in der entwicklungspolitischen Zusammenarbeit. Spezial 122. Bonn: BMZ.
- BMZ** (2005c): Mehr Wirkungen erzielen: Die Ausrichtung der deutschen Entwicklungszusammenarbeit auf die Millenniums-Entwicklungsziele. Die Umsetzung der Paris Declaration on Aid Effectiveness. Spezial 130. Bonn: BMZ.
- BMZ** (2008): Medienhandbuch Entwicklungspolitik 2008/2009. Bonn: BMZ.
- BMZ** (2011): Zehn Ziele für Bildung – BMZ Bildungsstrategie 2010–2013. URL (15.09.2011): [http://www.bmz.de/de/publikationen/reihen/strategiepapiere/Strategiepapier315\\_1\\_2012.pdf](http://www.bmz.de/de/publikationen/reihen/strategiepapiere/Strategiepapier315_1_2012.pdf)
- DCED** (2010): The DCED Standard for Measuring Achievements in Private Sector Development. Control Points and Compliance Criteria. Version V. DCED.
- DCED** (2011): Why have a Standard for measuring results? Progress and plans of the Donor Committee for Enterprise Development. DCED.
- Diekmann, Andreas** (2007): Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen. Reinbek bei Hamburg: Rowohlt.
- Erlor, Brigitte** (1985): Tödliche Hilfe. Bericht von meiner letzten Dienstreise in Sachen Entwicklungshilfe. Köln: Hayit Diskurs (14.).
- Georg, Walter** (2006): Berufsbildung in Entwicklungsländern, In: Arnold, R./Lipsmeier, A. (Hrsg.), Handbuch der Berufsbildung. Wiesbaden: Verlag für Sozialwissenschaften.
- GTZ** (2006): Handreichung zur Bearbeitung von AURA-Angeboten. Eschborn: GTZ.
- GTZ** (2007): The world of Words at GTZ. Eschborn: GTZ.
- GTZ** (2008): Results-based Monitoring – Guidelines for Technical Cooperation Projects and Programmes. Eschborn: GTZ.

- GTZ** (2010): Baselineerhebung, Ein Leitfaden zur Planung, Durchführung, Auswertung und Nutzung der Ergebnisse. Eschborn: GTZ.
- Haldemann, Theo** (2009): Social Info. Wörterbuch der Sozialpolitik. URL (13.11.09): <http://www.socialinfo.ch/cgi-bin/dicopossode/show.cfm?id=735>
- Honig, Michael-Sebastian/Betz, Tanja/Schreiber, Norbert** (2007): Bildung in Rheinland-Pfalz. Bildungschancen: Beteiligungsformen und Verlaufsmuster. Vorstudie zu einem Bildungsbericht in Rheinland-Pfalz. Trier: Zentrum für sozialpädagogische Forschung Trier.
- IHK** (2009): Das Geschäftsfeld Aus- und Weiterbildung. URL (13.11.2009): <http://www.bielefeld.ihk.de/german/main-navigation/bildung/geschaeftsfeld0/>
- Kevenhörster, Paul/Boom, Dirk van den** (2009): Entwicklungspolitik. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Kluve, Jochen/Boldemann, Hanka/Weidnitzer, Eva** (2010): Measuring employment effects of technical cooperation interventions – Some methodological guidelines, GTZ.
- Lahner, Jörg/Cisz, Katarzyna** (2004): Berufliche Bildung in der Entwicklungszusammenarbeit – Elemente einer Neuorientierung.
- Ergebnisse der Expertenkonferenz in Nordhausen** März 2004. Göttingen: Seminar für Handwerkswesen an der Universität Göttingen.
- Lutsyk, Tetyana** (2010): Jenseits der Beschäftigungsquote – eine Orientierung zur Messung von Beschäftigungsfähigkeit in Vorhaben der Beruflichen Bildung. GTZ.
- Meyer, Wolfgang** (2007): Messen: Indikatoren – Skalen – Indizes – Interpretationen. In: Stockmann, Reinhard (Hrsg.), Handbuch zur Evaluation. Eine praktische Handlungsanleitung, Münster u.a.: Waxmann.
- OECD** (2007): Bildung auf einen Blick 2007. OECD-Indikatoren. Bielefeld: Bertelsmann Verlag.
- Oxfam u.a.** (2007): Impact Measurement and Accountability in Emergencies: The Good Enough Guide. Parkstone: BEBC Publishing.
- PMA Projektmanagement Austria** (2008): pm baseline V 3.0 (Grundlage für Projektmanagement-zertifizierung gemäß der IPMA), Wien: pma.  
<http://www.p-m-a.at/Download-document/6-pm-baseline-3.0-Deutsch.html>.
- PMI Projektmanagementinstitute** (2008): A Guide to the Project Management Body of Knowledge (PMBOK® Guide) – Fourth Edition (auch in Deutsch erhältlich), Project Management Institute: Newton Square, PA (Kontakt: <http://www.pmi-berlin.org/>).

- Reich, Ursula** (2010): Enquete sur les indicateurs et le suivi relatifs au degré de l'intégration familial, socio-économique et sociale des jeunes. FAKT Consulting.
- Riedel, A.** (2004): Didaktik der beruflichen Bildung. Stuttgart: Steiner.
- Seeber, Susan** (2006): Benchmarking in der beruflichen Bildung, In: Rauner, Felix (Hrsg.), Handbuch Berufsbildungsforschung. Bielefeld: Bertelsmann Verlag.
- Seyfried, Erwin** (2007): Indicators for Quality in VET. To enhance European Cooperation. Luxembourg: Office for official Publications of the European Communities.
- Stockmann, Reinhard** (2006): Evaluation und Qualitätsentwicklung. Münster: Waxmann.
- Stockmann, Reinhard/Kohlmann, Uwe** (1998): Transferierbarkeit des Dualen Systems. Eine Evaluation dualer Ausbildungsprojekte in Entwicklungsländern. Eschborn, Berlin: GTZ.
- Tippelt, R.** (2009): Bildung in Entwicklungsländern und internationale Bildungsarbeit  
In: Tippelt, R./Schmidt, B. (Hrsg.), Handbuch Bildungsforschung. Wiesbaden: Verlag für Sozialwissenschaften.

## 8. Glossary

<b>Alternative hypothesis</b>	Assumptions regarding causes other than the measures undertaken by the intervention, which might lead to the same results.
<b>Attribution</b>	Attributing a change to one or several causes.
<b>Baseline</b>	A 'baseline study' is an 'analysis describing the situation prior to a development intervention, against which progress can be assessed or comparison made' (OECD 2009:23).
<b>Before-after-comparison</b>	Comparison of the results obtained at two different times of measurement. It does not cover the differences between different groups.
<b>Causality</b>	Relationship between cause and effect.
<b>Comparative analysis</b>	Comparison of at least two measurement results. The measurements can be taken in different groups (treatment and control group), at different points in time (before-after-measurement) or regarding different issues (measuring of attitudes).
<b>Comparison group</b>	Group of people who are as similar as possible to those in the intervention group. In contrast to a control group, comparison groups are established after the intervention, using either natural groups (e.g. parallel classes) or groups mathematically identified following assignment probabilities (e.g. propensity score matching).
<b>Control group</b>	Group of people who were not exposed to the intervention. Unlike a comparison group, the members of the control group and the experiment group are randomly assigned before the intervention.
<b>Counterfactual</b>	Condition that would have prevailed if no measures had been undertaken.
<b>Cross-sectional comparison</b>	Comparison of two or more group results at a single time of measurement. Time-related changes are not recorded.
<b>Decent Work</b>	The notion of decent work was developed by the International Labour Organization (ILO) and is based on four pillars: (1) Promotion of employment: Work and a fair income, sufficient to meet their basic needs, allow a life in dignity. (2) Social security: Decent work means access to social protection for workers and their families, such as benefits in the case of sickness, unemployment, incapacity, work injury, maternity or loss of a main income earner. (3) Compliance with core labour standards: Core labour standards eliminate forced labour, prevent child labour and discrimination in employment and occupation and ensure freedom of association and the right to collective bargaining. (4) Social dialogue: Social dialogue links the pillars with each other and promotes the equal dialogue and cooperation between representatives of employers, workers and governments.
<b>Diffusion</b>	The spreading of the results achieved by a project at a certain point beyond its immediate sphere of influence.
<b>Disturbance</b>	Factors influencing the cause-effect-relationship of a development intervention or its measurement.

<b>Employment</b>	Graduates practice an occupation, the work requirements of which correspond with the qualification attained during a reformed vocational training programme and which receives appropriate payment.
<b>Experiment</b>	A set up for measurement and evaluation which seeks to isolate the connection between one cause and one result and to control all other influences. One core element of an experimental design is the random assignment of subjects to two groups, of which only one is subjected to the influence to be observed by the experiment. Measuring twice – before and after the intervention – and comparing the two groups makes it possible to analyse and evaluate the cause-and-effect relationship.
<b>Experimental group</b>	The group subjected to an intervention/treatment within an experiment (also called treatment group).
<b>Expert interviews</b>	Non-standardised or semi-standardised survey among people who can provide profound insight into a certain topic due to their experience and/or knowledge. Usually these discussions are guided by a set of questions but do not use pre-established choices of answers (open interview) to allow the interviewees to use their expertise to answer the questions more flexibly.
<b>Focus group discussion</b>	Qualitative survey method where a group of experts give their opinion on a certain issue in the course of a moderated discussion. Usually, the aim is to achieve a joint result (not necessarily consensual).
<b>Guided interview</b>	See expert interview.
<b>Indicator</b>	Parameter used to depict a certain, often complex issue that cannot be measured directly.
<b>Interventions</b>	Consciously undertaken manipulations of a social sphere; measures undertaken to change a certain condition.
<b>Key indicators</b>	Centrally important parameters in a sector, which seek to express results that are relevant to all projects in the sector.
<b>Longitudinal study</b>	Several measurements (at least three) that are taken at regular time intervals.
<b>Market-Player-Opinion</b>	Assessment of the market situation and of the behaviour of market players by third parties who have studied the market thoroughly and have a good comprehensive view of it.
<b>Measuring</b>	Comparing an object of study with another, normed object on the basis of a shared dimension with established categories.
<b>Monitoring</b>	Regularly repeated measurement of indicators, which provide information on a condition development process relevant for the project's progress.
<b>Panel study</b>	Gathering of data from the same people at regular intervals. This allows analysis of individual developments and paths.

<b>Process indicators</b>	Parameter depicting the relevant aspects of the implementation of a measure (inputs, activities, outputs and use of outputs).
<b>Qualitative design</b>	Methodological procedure for collecting data geared towards understanding processes and behaviour in an open procedure.
<b>Quasi-experimental design</b>	A set up for measurement and evaluation, which due to external conditions only approximately fulfils the requirements of an experiment (for instance when during field research it is not possible to randomly assign the participants of the intervention and it becomes necessary to use already existing comparison groups).
<b>Randomisation</b>	Random assignment of individuals to the experimental and control groups to control the influence of person-related properties and attitudes on the experiment.
<b>Reliability</b>	The replicability of a measurement, i.e. repeating the measurement under equal frame conditions always provides the same result.
<b>Results</b>	The changes caused by a development intervention. They can be intended or unintended, positive or negative. Results occur during the entire course of a project or programme. They do not only affect the previously defined target groups but also partners and mediators and may occur in many areas.
<b>Results chain</b>	A sequence of consecutive cause-and-effect connections used when planning a project or programme. Ideally, inputs lead to activities, which then achieve outputs. By making use of the outputs, the outcome is brought about, which in turn contributes to reach overarching development goals (impact).
<b>Results expectation</b>	An assumption on the effects to be achieved with an intervention/measure (assumption of causality). These expectations may refer to the immediate environment of the project and the results produced there (direct results/outcome) or to sequences of action beyond the direct environment of the project and affecting overarching development goals (indirect results/impact).
<b>Results hypothesis</b>	Hypothetical assumption regarding the relationship between a cause and a result.
<b>Results indicators</b>	Parameters depicting results. They do not necessarily measure the results directly but may indicate the attainment of a certain stage of development and thus show that the results (that have not been achieved yet) are 'on the way' to be achieved.
<b>Scope</b>	Share of the participants of certain vocational training programmes with regard to the entire target group.
<b>Scope indicator</b>	Parameter used to depict the spreading of measures and/or results within the target groups.
<b>SMART</b>	Acronym for the quality criteria of indicators (specific, measurable, achievable, relevant, time-bound).
<b>Sustainability</b>	Durability of results and effects which could be achieved by the measures of an intervention.
<b>Target</b>	Numerical value of an indicator to be achieved at the end of an intervention.

<b>Target-performance analysis</b>	Ratio between the achieved and intended values of a certain performance indicator.
<b>theoretical construct</b>	A complex (social) phenomenon containing various dimensions. The terms 'happiness', 'satisfaction', 'peace', 'democracy', 'quality of teaching', etc. are some examples. In our example this particularly applies to the term 'employability'.
<b>Threshold</b>	Numerical value, which must (minimum) or must not (maximum) be exceeded. It is usually applied in technical contexts (e.g. vehicle exhaust emissions testing).
<b>Treatment group</b>	The group subjected to an intervention/treatment within an experiment (also called experimental group).
<b>Trend survey</b>	Analysis using data gathered at regular intervals. As opposed to the panel study, developments and paths can only be assessed at an aggregate level (e.g. the cohort of graduates of vocational training courses in a particular occupational area), because the data gathered at different points in time are not connected to each other and cannot be assigned to individual persons.
<b>Validity</b>	Soundness of a measurement, i.e. the measuring instrument only measures the changes to the measurement dimension and is not influenced by other factors.





Published by

Deutsche Gesellschaft für  
Internationale Zusammenarbeit (GIZ) GmbH

Registered offices  
Bonn and Eschborn, Germany

Sector programme 'Technical and vocational education and training'

Friedrich-Ebert-Allee 40	Dag-Hammarskjöld-Weg 1–5
53113 Bonn / Germany	65760 Eschborn / Germany
Phone: +49 228 44 60 -0	Phone: +49 6196 79 -0
Fax: +49 228 44 60 -1766	Fax: +49 6196 79 -1115

[berufsbildung-arbeitsmarkt@giz.de](mailto:berufsbildung-arbeitsmarkt@giz.de)

Writer

Dr. Wolfgang Meyer  
Stefan Thomas

Design

andreas korn visuelle kommunikation, Bad Homburg

Printed by

Aksoy Print & Projektmanagement, Eppelheim  
Printed on FSC-certified paper

Photographs

Cover: © GIZ/Nicole Herzog

As at December 2011

GIZ is responsible for the content of this publication

On behalf of

Federal Ministry for Economic Cooperation and Development,  
Division Education

BMZ Bonn	BMZ Berlin / im Europahaus
Dahlmannstraße 4	Stresemannstraße 94
53113 Bonn / Germany	10963 Berlin / Germany
Phone: +49 228 99 535 -0	Phone: +49 30 18 535 -0
Fax: +49 228 99 535 -3500	Fax: +49 30 18 535 -2501

[poststelle@bmz.bund.de](mailto:poststelle@bmz.bund.de)  
[www.bmz.de](http://www.bmz.de)