

Part A

Introduction

Purpose and use of the common questions with criteria and indicators

1. BACKGROUND PURPOSE AND CONTENTS

Regulation 1750/99¹ (Article 42(2)) requires the Commission in consultation with the Member States to define common evaluation questions with associated criteria and indicators for the rural development programmes 2000-2006, in particular concerning the *mid-term* and *ex-post* evaluations. The consultations have in particular involved the STAR Committee (Committee on Agricultural Structures and Rural Development), and the resulting common elements for the *mid-term* and *ex-post* evaluations are in Part B. **Box 1.1** provides an overview of the documentation concerning these common elements as well as previous and planned documents concerning evaluation of rural development programmes 2000-2006.

The present introduction (Part A) describes the use of the set of common evaluation questions, judgement criteria and indicators and it reviews the links between these common elements and other aspects of the *mid-term* and *ex-post* evaluation at national/regional level (e.g., *the quantified objectives corresponding to the strategy of individual programmes, issues concerning the collection and analysis of data, context analysis etc.*). All references to articles in this text relate to Regulation 1750/99 and the evaluation terms are explained in Annex IV of the ‘Guidelines’ mentioned in Box 1.1.

Box 1.1 Documents concerning the common elements and previous documents about evaluation of rural development programmes 2000-2006

Documentation concerning the common elements:

- Part B: THE SET OF COMMON EVALUATION QUESTIONS WITH CRITERIA AND INDICATORS FOR RURAL DEVELOPMENT 2000-2006:
 - elements that must be used whenever possible; non-application must be substantiated as required in Article 44(1)
- Part C: ECONOMIC TERMINOLOGY (auxiliary document):
 - vocabulary and definitions used in the documents
- Part D: EXPLANATORY SHEETS (auxiliary document):
 - support for the independent evaluator to avoid misunderstandings concerning the common elements and to improve the quality of evaluations
 - complementary information (clarifications, suggestions ...) of a non-constraining nature:
 - key elements of the intervention logic concerning each question
 - calculating procedures, suggestions for breakdowns of the indicators...
 - typologies (existing) for structuring data collection, analysis, reporting

Previous document:

- EVALUATION OF RURAL DEVELOPMENT PROGRAMMES 2000-2006 SUPPORTED FROM THE EUROPEAN AGRICULTURAL GUIDANCE AND GUARANTEE FUND – GUIDELINES (referred to as ‘Guidelines’ in the present document)
 - general information: linking basic evaluation concepts to the requirements of Regulation 1750/99 (e.g., structure of the evaluation process, quality and recognised evaluation practice, independence ...)
 - administrative matters: details about terms of reference, co-funding, format and deadlines for reporting...

Planned document(s):

- Guidelines concerning the approach to the *mid-term* and *ex-post* evaluations (c.f., Articles 44(2) and 42(2) about a differentiated approach)

¹ Commission Regulation (European Commission) No 1750/99 of 23.07.1999 laying down the detailed rules for the application of Council Regulation No 1257/99 on support of rural development from the European Agricultural Guidance and Guarantee Fund (EAGGF), OJ L 214, 13.08.1999, p.31

2. GENERAL EVALUATION STRATEGY FOR THE ENTIRE PROGRAMME PERIOD

2.1. GENERAL OUTLINE OF THE STRATEGY

The strategy for the evaluation of rural development must assure a global view of the entire evaluation process from 2000 to 2009, both enabling regional/national authorities to plan and organise evaluations of high quality in complementarity to monitoring control and audit, and enabling the Commission to elaborate an EU-level evaluation synthesis concerning rural development. It must also assure evaluations with sufficient emphasises on analysis, not just description of the financial inputs and the direct programme outputs.

Regulation 1750/99 provides a framework for the evaluation strategy. Firstly, it defines a permanent partnership with differentiated responsibility between national/regional authorities and the Commission. Secondly, it identifies the elements that all evaluations must apply. Thirdly, it combines a coherent evaluation process across programmes with flexibility by focusing on effects that can be expected across many programmes due to the means and objectives of Regulation 1257/99², while only co-ordinating the methodological aspects through a quality requirement. Fourthly, it foresees a limited differentiation between the mid-term and ex-post evaluation stages in how to apply the elements (questions, criteria, indicators...) of the evaluation (Article 44(2)). **Box 2.1** provides more detail about the elements of this

Box 2.1 Elements of the strategy for mid-term and ex-post evaluation of rural development programmes

Key elements of the evaluation strategy providing a global view of the process 2000-2009:

1. Permanent partnership between authorities responsible for managing rural development programmes and the Commission:
 - the Commission defines the common elements (in consultation with Member States) (Article 42(2))
 - the regional/national level is responsible for performing the evaluations (in consultation with the Commission) (Article 43(2), 44(3))
2. Elements that all evaluations must apply:
 - programme-specific: complementary questions with criteria & indicators (Article 44(1))
 - common elements: common questions with criteria & indicators (focusing on effects rather than measures) (Articles 42(2), 44(1))
3. Differentiated extent of harmonisation for various parts of the evaluation process
 - co-ordinated approach to evaluation of such results/impacts that can be expected across a large number of programmes
 - generally no methodological harmonisation (data collection, analysis), but a quality requirement (Articles 40, 42(1), 44(4) and 45(3))
 - some harmonisation of reporting (so the Commission can elaborate a Community level synthesis) (Article 45(3))
4. Limited differentiation between the *mid-term* and *ex-post* evaluations in how to apply common elements (in particular the indicators) (Article 44(2))

strategy.

² Council regulation (EC) No 1257/1999 of 17.05.1999 on support of rural development from the European Agricultural Guidance and Guarantee Fund (EAGGF) and amending and repealing certain Regulations, OJ L 160, 26.06.1999, p.80

2.2. THE COMMON ELEMENTS FOR ALL INDIVIDUAL PROGRAMMES

2.2.1. Using the common elements

All common evaluation questions are accompanied by judgement criteria and indicators, which are integral parts of the questions (Article 42(2)). The independent evaluator must hence apply all these common elements together whenever they are relevant to the eligible actions and the context of a particular programme. Otherwise, an underpinned explanation is necessary (Article 44(1)).

A question with its criteria and indicators is inappropriate where a whole chapter of Regulation 1257/99 with all its measures is not implemented or if a separate sub-chapter, such as ‘areas with environmental restrictions’ in Chapter V, is not implemented. However, such automatic exemption does not apply if:

- Only part of the actions/measures of a chapter are not implemented. This is because the evaluation strategy is an effect-wise evaluation (results, impacts), not measure wise³ (outputs) and many results and impacts derive from several types of actions (this situation is closely connected to the following case).
 - It is sometimes possible that a particular question is pertinent, but not all its associated criteria and indicators are applicable. This is discussed in sections 2.2.3 and 2.2.4.
- There are distinct positive collateral effects, i.e., programme effects that respond to the objectives of Regulation 1257/99 without being explicit objectives of the particular rural development programme or side-effects of investments with another aim (*e.g., investments in new buildings or equipment for economic reasons will often cause environmental benefits as well, because new technology is likely to be more environmentally benign than the one it replaces*). Such collateral effects ought also to be considered, because of their relevancy to the regulation and because routine omission would lead to a systematic under-estimation of the benefits of the rural development policy. Obviously, much less effort should be put into the examination of collateral effects than of the key objectives of the programme (see Section 2.2.4 concerning proportionality).
- The intended effects are (yet/apparently) insignificant or difficult to measure (c.f., Section 2.2.4).

Box 22 Cases where the range of effects examined by a common question seems to exceed the effects, likely to arise from the means and objectives of a given programme

For the cases (a) - (e) below the effects examined by a common question (with criteria and indicators) may seem to exceed the effects likely to arise from the means and objectives of a given programme. For some of these cases it will be absolutely justified (c.f., Article 44(1)) not to use a particular common question but in other cases most likely not, and the need for a well-underpinned justification of a possible non-application of certain common elements varies accordingly:

Case (a): a whole Chapter of Regulation 1257/99 is not implemented:

non-application normally unproblematic → little explanation required

Case (b): a whole sub-chapter of Regulation 1257/99 is not implemented:

(non-application normally unproblematic) → little explanation required

Case (c): part of the actions/measures within a Chapter of Regulation 1257/99 are not implemented:

non-application not automatically justified, but a certain amount of proportionality may be appropriate, c.f. Sections 2.2.4 and 3.2 → underpinned explanation required

Case (d): significant collateral effects are present (i.e., effects covered by the means and objectives of 1257/99 but not included in the objectives of the individual programme):

non-application most often not justified, but proportionality is appropriate, c.f. Sections 2.2.4 and 3.2 → underpinned explanation required

Case (e): no perceivable effects (yet or apparently):

non-application not automatically justified, but the use of the indicator may be modified in some cases or programme amendments may be needed (c.f., Section 2.2.4 and Box 2.4) → underpinned explanation required

Example concerning cases (c) and (d):

Water quality depends on such a variety of eligible actions that it would be difficult to imagine an agri-environmental programme for which the question about water quality will not at all be relevant, and as Chapter VI is compulsory, it is difficult to imagine an evaluation where this question will be entirely inappropriate. Therefore, the possible non-use of a question of this type, requires a well underpinned justification. However, specifically some of the indicators (the most complicated ones) for the water quality question are marked with an asterisk, signifying that this part of the common elements needs only be used for programmes with a certain focus on water protection (e.g., where relevant actions are applied to catchment areas predominately influenced by farming and forestry).

Box 2.2 provides an overview of situations where the common evaluation questions examine a wider field of effects than covered by the means and objectives of a particular rural development programme (the opposite issue, i.e., where the specific programme contains elements going beyond the common questions is dealt with in Section 2.3.1).

2.2.2. Common evaluation questions

The common questions (occasionally sub-questions dealing with aspects of a question) concern matters relevant at EU-level (Article 44(1)). They examine programme effects (i.e., results, impacts) that can be expected thanks to the means and objectives of Regulation 1257/99 (only effects expected within a significant number of programmes).

Most questions focus on the effectiveness of the programmes, but they also investigate other key aspects of evaluation where this is appropriate and feasible at EU-level (see Box 3.1 of the Guidelines about aspects such as relevance, effectiveness, efficiency, utility and sustainability of the results).

The majority of questions are chapter-specific, i.e., they concern effects arising from the eligible measures within each of the nine chapters of Regulation 1257/99. Most chapters embrace several individual measures/actions, so the questions will most often be relevant even if some eligible measures of a chapter are not applied.

Six questions are cross-cutting, examining effects that can be expected to arise from across the entire programme. This means that (a) they examine whether the programming approach has led to added value compared to a pure juxtaposition of the individual chapters/measures, (b) they look at benefits to both direct and indirect beneficiaries (the chapter-specific questions relating to chapters VI and IX partly do so too) and (c) they regard significant collateral effects (i.e., side-effects from the point of view of a particular programme, but covered by the EU-level objectives of Regulation 1257/99). Finally, they provide a summary concerning the main effects of the programme in relation to the means and objectives of Regulation 1257/99. It is expected that all cross-cutting questions will be applicable to almost all individual programmes.

The answer to each evaluation question should be based on the common criteria and indicators and any other relevant information about the impacts of the relevant implemented actions. It should not just be a mechanical application of the indicator(s) through their criterion; it must present and discuss the evidence for each judgement criterion, in particular the common programme indicators. It should also consider any programme specific complements (e.g., additional programme indicators), the context (e.g., *exogenous factors*) and relevant evaluation concepts such as dead-weight, etc. The examination of the context may for example help determine net effects and situations where the measures have worked particularly well and other situations where this has not been the case. It may be relevant to differentiate certain answers to sectors or zones within the programme (e.g., *break-downs to type of holding, age of beneficiary, pastoral zones...*). This can improve the answers and enable comparison of different natural and socio-economic situations within and among programmes.

2.2.3. Judgement criteria

The criteria are decisive factors for judging whether a particular evaluation question can be positively replied to (see Guidelines Section 4.3).

Some questions include more than one criterion, normally (a) because the question is broadly formulated (e.g., *comprising both erosion and soil contamination*) or (b) because several stages in the transformation of the inputs to the intended effects are examined (e.g., *where the effect needs long time in order to materialise or is difficult to measure*).

This means that the criteria are an inseparable part of the common questions and all criteria must be addressed, if they are in any way relevant to the circumstances of the programme. The use of the criteria is a very important part of the quality requirement of Regulation 1750/99 and for the purpose of the EU-level evaluation synthesis, it is essential to know which criteria did contribute to a positive answer.

Most common criteria are formulated so as to make them generally applicable across a large number of individual programmes operated in dissimilar conditions and with different means and objectives. This means that, most often, a precise target level needs to be added to the criteria/indicator combination. This must be done at the level of the individual programme (see Section 3.1.2 and the Guidelines Section 4.3). The attached explanatory sheets contain suggestions and examples of such target levels for each indicator.

A criterion can seldom be abandoned if the question to which it is associated is pertinent, but certain cases are possible, *for example, criterion 2 of Question IX.4 ('agricultural production potential has been protected/restored regarding natural hazards'), which depend on a limited number of eligible action that may not have been implemented within a particular programme.*

2.2.4. Programme indicators

The purpose of the common indicators is to indicate whether or not a particular judgement criterion is fulfilled. In general, they measure results or impacts, although several of them relate to outputs in order to ascertain that at least the initial steps of the delivery mechanism for the intended/ expected effects have been engaged by the programme. In the latter case, it is normally types of output (*e.g., “area subject to input-reducing actions thanks to agreement – of which (a)...(b)...”*) rather than the specific actions themselves (*e.g., organic farming, integrated farming...*) that serve as indicator; this helps avoid that the feasibility of using the indicator varies excessively between programmes. Some common indicators are for practical reasons (costs of data collection, measurability ...) more restricted in scope than the criterion they relate to, so it will sometimes be useful to complement them with additional indicators at the level of individual programmes (Section 2.3.2).

Most indicators are quantitative, *inter alia* because they are easier to aggregate at the EU-level, but 'trends' or other qualitative approaches (*e.g., simply asking for whatever available 'evidence'*) are also used in several cases.

Even though the indicators are an integral part of the common questions (Article 42(2)), more flexibility is justified for the use of the indicators than for the questions and criteria. This is because in certain programmes, a small number of the common evaluation indicators may be inapt for indicating the effects they are intended for even if they are formulated so as to make them generally applicable. **Box 2.3** provides an overview of how such flexibility can be achieved, i.e., through (a) proportionality in the effort of quantifying the indicators, (b) replacing an indicator with another that is unmistakably better in the given situation or (c) entirely abandoning an indicator. It should be noted that this flexibility is over and above the indicators that have been entirely abandoned together with their question, for reasons (a) and (b) in **Box 2.2**.

Replacing or abandoning an indicator is acceptable for a limited number of indicators within a programme, but not as a systematic approach (otherwise, the Commission cannot establish a good

Box 2.3 Flexibility in the use of the common indicators

Flexibility in the use of an indicator may involve:

- Proportionality in the effort of quantifying an indicator (effort compared to utility):
 - smaller samples, case studies...
 - using coefficients for transforming monitoring data to effects
 - calculating the common indicator only for key combinations of actions/measures versus socio-economic/natural context
 - using less onerous data, provided (a) the indicator lend itself to this and (b) it does not concern a key objective of the programme (*e.g., using kilogram of commercial substance instead of kilogram of active ingredient for quantifying indicator 1.2 of Question VI.2.A, for programmes where this indicator does not measure a key objective*)
- Replacing an indicator with another, which is unmistakably better in the given context:
 - acceptable for a limited number of indicators, but not systematically
 - a justification is needed for disregarding the common indicator
- Entirely abandoning an indicator, if it depends on one/few actions that have not been implemented: (*e.g., indicator 1.1 ('agricultural improvement') of Question IX.4 which depends on the application of a limited number of measures such as the setting up of farm services or reparing of farmland; or indicator 1.2 ('holdings connected to associations of forest holders or similar') of Question VIII.2.A which mainly depend on one type of action*).
 - acceptable for a limited number of indicators, but not systematically
 - a justification is needed for disregarding the common indicator

Community level synthesis) and a justification is needed in such cases. A perceived need for flexibility may only be apparent. **Box 2.4** provides a tentative overview of situations where there is a real, apparent or temporary lack of effect, or where the lack of measurable effect is due to a serious lack of strategy (concentration) within the programme. This overview clearly illustrates that it often will be inappropriate just to abandon all attempts of quantifying the common indicator, because the difficulty encountered may point at severe deficiencies in the perception or management of a programme which need to be investigated.

Indicators should in general be provided on the same geographical scale as the rural development plan. However, some actions/measures may not apply throughout the entire zone covered by the programme (e.g., *less-favoured areas or areas with environmental restrictions*) or be differently targeting within the zone covered by the programme. In such cases the indicators should be broken down in order better to determine the effects of and influence on intra-regional or inter-sectoral disparities (the explanatory sheets suggest certain breakdowns and lists existing typologies for each chapter for this purpose).

Box 2.4 Cases where it is difficult to indicate effects

Indicators may seem difficult to apply or insensitive when the intended effects are (yet or apparently) insignificant or difficult to measure because of the reasons listed under the cases (a)-(e). The degree of flexibility for using the indicator and the action to be taken by the evaluator vary for these types of cases:

Case (a): The effects are likely to arrive but have not yet become manifest (e.g., at the *mid-term stage*):

→Regulation (Article 44(2)) foresees flexibility for this problem at the *mid-term* stage (as explained in Section 3.1.3))

Case (b): Difficulties in measuring the effects, either technical difficulties or because of the cost involved (e.g., because the programme does not focus much on the related objective):

→Flexibility is justified for a limited number of indicators within a programme, but not as a general approach, c.f., Box 2.2 Case (e)

Case (c): Unfavourable contextual/exogenous factors work against the programme effects:

→It is important to analyse such situations to find out whether:

- net effect is positive but outweighed by erratic negative exogenous factors (i.e., the situation of the beneficiaries improve compared to that of non-beneficiaries)
- actions do not work because they are applied in an erratic adverse natural/socio-economic situation (i.e., the evaluator should identify situations where it does work and recommend how to adjust programme or procedures)

Case (d): Shortcomings of the programme [*inadequate intervention logic, lack of “concentration” (i.e., lack of strategy, e.g., where insignificant amounts of aid has been distributed to a large population)*]:

→Such cases must lead to a clear recommendation from the independent evaluator about improving the strategy of the programme at the *mid-term* stage; not just to abandoning all attempts of identifying the effects and passing tacitly over the problem, c.f., the quality requirement of Regulation 1750/99.

Case (e): Inadequate programme management (e.g., because the beneficiaries most in need and with the highest potential of obtaining results are not selected):

→Same conclusion as for case (d)

It should be noted that the prescribed breakdowns of certain indicators (“of which ...”) do not add up to 100%. The sum may be more than 100 because different typologies are mixed or less than 100% because it is not feasible or worthwhile, at the EU-level, to predict all possible elements of the breakdown.

The relation between the principal types of indicators (monitoring, evaluation, context) is explained in Sections 3.1.2 (Box 3.2) and 3.3.2.

2.3. THE PROGRAMME SPECIFIC ELEMENTS

2.3.1. Purpose and use of the programme specific elements (strategy/ objectives of programme)

The detailed means and objectives of the individual programmes will deviate from those in Regulation 1257/99. They must fit into the EU-level objectives and will also be linked to general national/regional objectives in the same domain. The common elements (questions, criteria and indicators) can therefore not cover all aspects of all programmes and the evaluation strategy in Regulation 1750/99 (Articles 44(1) and 45(3)) foresees that programme specific evaluation matters should be added in the terms of reference and/or by the independent evaluator (the opposite problem, where the common evaluation questions cover issues not included in the particular programme is discussed in Sections 2.2.1 and 2.2.4 above).

The individual evaluations must hence cover the global strategy and the objectives of the specific programme. The evaluator must therefore also evaluate the programme against its own quantified objectives at the operational (outputs), specific (results) and global (impacts) level. Such quantification should to a large extent already have taken place at the planning and *ex-ante* stage as required in point 6(1) 3rd indent of the annex to Regulation 1750/99.

Another reason for including programme specific elements is that the common evaluation questions are strongly focused on the intended programme effects. Unexpected, possibly negative, effects are by their nature disparate and difficult to forecast so the common evaluation questions do in general not inquire about such effects (cross-cutting questions 5 and 6 being the most notable exceptions).

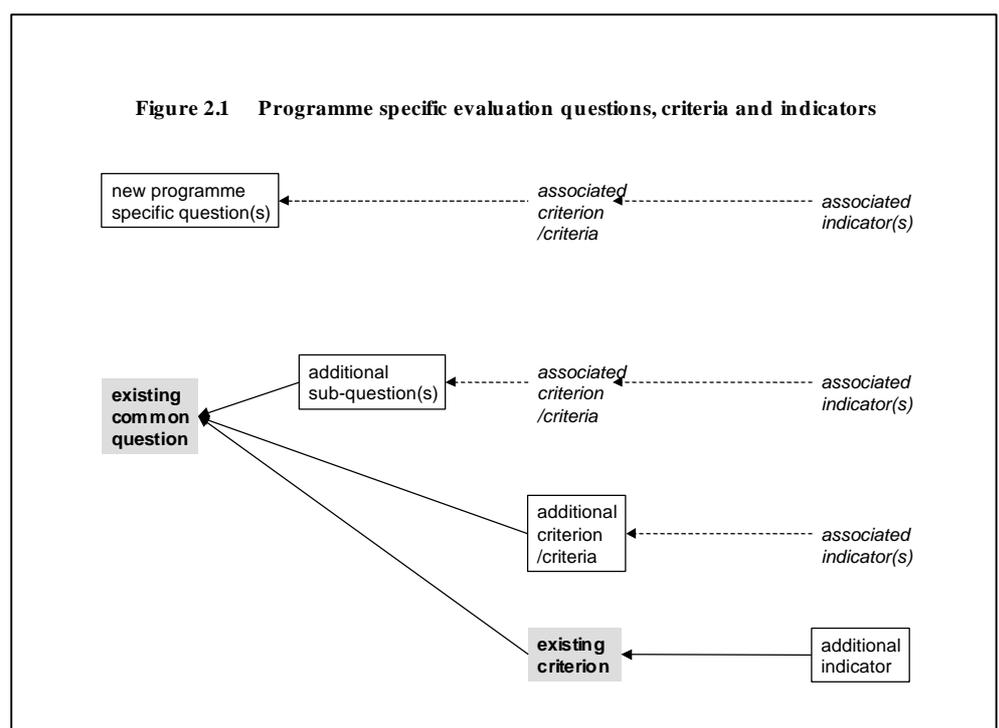
The programme specific evaluation must also look more closely at efficiency (Article 44(2)) to the extent that this is not covered by the common evaluation question, c.f., Section 2.2.2. Finally target level must be added to most criteria as explained in Section 3.1.2 below.

2.3.2. Additional questions, criteria and indicators

The common strategy foresees no harmonisation of the programme specific elements, except that the regulation assumes that the approach of questions, criteria and indicators will be used (Articles 44(1) and 45(3)) and the usual quality requirement applies.

The use of programme specific evaluation questions, criteria and indicators together with the common ones is illustrated in **Figure 2.1** (and in Figure 4.2 of the Guidelines). The box illustrates the types of elements that can be added:

- entirely new questions (with criteria and indicators) covering effects not dealt with by the common questions
- new sub-questions (with criteria and indicators) covering effects not fully dealt with by the common questions;
- new criteria (with indicators) covering judgements aspects not dealt with by the common criteria
- new indicators covering information not dealt with by the common indicators



3. IMPLEMENTATION OF THE COMMON EVALUATION STRATEGY

3.1. STAGES OF THE EVALUATION PROCESS

3.1.1. Outline of the steps in the evaluation process

The process of defining the programme-specific elements presupposes a foregoing structuring process in order to identify the impacts that can be expected from a particular programme so that appropriate indicators can be identified and subsequently quantified through various forms of collection of information. This and all other relevant information must be analysed (including likely cause/effect relations and context influences; cross-validations...) and judgements be formulated in the form of answers to the common questions and other evaluation aspects. The final stage is the reporting at various levels (regional, possibly national and EU-level).

3.1.2. Preparing comparisons: baselines/target level

The structuring process for the common elements need to be completed at the level of the individual programmes because many of the common judgement criteria are formulated so as to make them applicable across a large number of programmes (c.f., Section 2.2.3) so they are not fully operational without a description of how exactly to perform the comparison they prescribe, and without a quantification of the baseline (*the remark about flexibility in the formulation of the common criteria does, in reverse, mean that separate target levels may be superfluous for the programme specific criteria, provided that they are formulated with enough precision for each indicator*).

It is an important task for the authorities responsible for the programmes to ensure that both these tasks take place at an early stage, possibly in interaction with the independent *mid-term* evaluator where he has been identified already, and the latter must check that all targets have been set in a sufficiently clear manner.

Comparisons can be temporal (before/after), counterfactual (policy on/off) and normative. These situations are outlined in **Box 3.1** and discussed in Section 4 of the Guidelines. These types of comparison have varying capacity to identify the net effects of the programme and lead to different resource requirements for the evaluation. The comparison with the counterfactual situation has the advantage of removing exogenous factors. This situation can, for example, be established through comparison with non-assisted peers, which in practice sometimes simply can be the average of the population (including the assisted population). However, this simplification will obviously not work if the assisted group forms a large proportion of the potentially eligible group (*e.g., for less favoured areas*), or where only few potential beneficiaries are genuinely comparable (*e.g., for processing and marketing*). The temporal comparison will obviously be difficult where a large part of the population has received more or less similar aid for a long time, *e.g.*,

Box 3.1 Main types of comparison

Three main types of comparison can be undertaken:

- **Temporal:** (the situation before/after for the beneficiaries)
The comparison requires information about the baselines, i.e., level of the indicator before an individual's or entity's participation in the programme. This information should normally be collected at an early stage of the programme. The baseline may sometimes be derived from the *ex-ante* evaluation (*e.g.*, collected for analysis). Purely temporal comparisons are sensitive to exogenous factors so it is difficult to isolate the net effect of the programme.
- **Counterfactual situation:** (the situation that would have occurred in the absence of the assistance)
The comparison require information about both beneficiaries and non-beneficiaries, *e.g.*, through sampling of information/trends for non-assisted peers models, ... Peer groups can often be approximated by using sectoral/regional statistics providing averages for entire (sub)populations provided that the ratio of beneficiaries to the comparison-population is sufficiently low.
This type of comparison can remove the effect of exogenous factors so that the net effect of the programme can be determined.
- **Norm:** (benchmark, best practice)
The comparison requires information about a norm, the best practice achieved previously or in other regions etc.

through the previous Objective 5a or Objective 5b programmes.

The suggestions for target levels presented in the Explanatory Sheets are non-compulsory recommendations and examples.

The independent evaluator must also examine the programme's performance compared to its quantified operational, specific and global objectives (corresponding to respectively outputs, results and impacts). The target levels will often not be identical to the quantified objectives of an individual programme; instead they complement the criterion/indicator-combinations in the particular situation applying to the programme, in order to enable a judgement about the success of certain aspects of the programme.

3.1.3. Differentiation between mid-term and ex-post evaluations

Right from the start of the programme, the evaluation strategy integrates the *mid-term* and *ex-post* evaluations in a unified strategy by focusing on the effects the evaluation must examine at the end of the programming cycle.

However, the *mid-term* and *ex-post* evaluations differ in terms of what is feasible (c.f., the different situations outlined in **Box 2.2**) and what is most relevant. This is described in Article 44(2) which explains that the *mid-term* evaluation – while answering the evaluation questions - will be more focused on initial achievements than the *ex-post* evaluation, while the latter will look more on impacts. A differentiated use of some common evaluation questions, criteria and indicators may hence be necessary and some answers may be partial at *mid-term* evaluation stage.

So, where the intended effects cannot yet be fully measured at the *mid-term* stage (i.e., indicators cannot be quantified), the independent evaluator must:

- (a) ensure that this is not due to a badly conceived or badly managed programme (e.g., using methods suggested in point (b) just below)
- (b) use proxy- methodologies, for example by verifying the extent to which the actions constituting the initial steps of the delivery mechanisms for the intended/expected effects have been engaged by the programme
- (c) if the effects cannot yet be expected to have materialised, it must be ensured that the necessary procedures in terms of monitoring and evaluation are in place for fully answering the questions at the *ex-post* stage

Box 3.2 Monitoring versus evaluation

- Monitoring examines the delivery of programme outputs (goods, services) to the beneficiaries thanks to the inputs (financial, administrative). It is a continuous process, carried out during the entire programme period, with the intention of immediately correcting any deviation from operational objectives.
- Evaluation examines particular results/impacts in depth at discrete points in the life cycle of a programme (*mid-term, ex-post*) in order to answer the evaluation questions. The information is only collected a couple of times during the programme (samples, case studies ...). However, evaluation also makes use of pure monitoring indicators in order to:
 - verify the initial step in the chain of causality (several indicators are typologies of outputs, e.g., in Chapter VI).
 - scale up impacts that have been identified through analysis of a limited sample (e.g., *coefficients linking outputs to effects*)
 - estimate efficiency.-----oo000oo-----
- It is not cost-efficient to collect all the information for evaluation (concerning results, impacts) through the monitoring system because monitoring indicators (concerning inputs and outputs) are collected (a) more frequently and (b) more intensively (for all beneficiaries) than the ad hoc collection of information for evaluation.
- Monitoring indicators are insensitive to:
 - effects materialising at a later stage.
 - effects materialising among indirect beneficiaries.
 - site-specific effects (one unit of output has different effects in different natural and socio-economic situations).
 - effects deriving from several outputs (or contrarily one type of output simultaneously contributes to several types of impacts)-----oo000oo-----
- Ad hoc information exclusively necessary for the evaluation process can be collected by the independent evaluator (samples, interviews, case studies ...) but the evaluator can in no case take over the regular monitoring (no co-financing is foreseen for monitoring).

Several common questions do already contain indicators that are suitable for the situation in point (b) above. This is especially commonplace in Chapter VI where many questions contain indicators situated at different points along the chain transforming the inputs to outputs, results and finally to impacts for the direct beneficiaries or the rural community. Proxy-methodologies may also involve the use of trends or qualitative information.

The Commission may, in due time before the *mid-term* evaluation, presents further guidance about the differentiated approach to the *mid-term* and *ex-post* evaluations in line with Articles 44(2) and 42(2).

3.2. INFORMATION SOURCES FOR EVALUATION

No particular methodology is prescribed for the data collection except the usual quality requirement of the regulation. The collection of primary and secondary information is described in detail in Section 5.2 of the Guidelines.

Much information can be supplied by the monitoring system and this is sufficient for quantifying certain indicators. In other cases, monitoring data can be transformed to impact indicators for all relevant beneficiaries, e.g., through coefficients established for smaller entities of beneficiaries within a typology. **Box 3.2** explains this and other aspects of monitoring in relation to evaluation indicators. The document about common monitoring indicators for rural development programmes (Working Document VI/12006/00) includes mainly input and output indicators.

Evaluation will always necessitate the ad hoc collection of a certain amount of quantitative data and qualitative information through interviews, samples, case studies, etc. This is part of the co-financed activity of the evaluator, but it can in no case be extended to include the administration's task of collecting ordinary monitoring information about inputs and outputs on a recurring basis.

A certain measure of proportionality (size of samples, case studies, use of coefficients from literature) may in certain cases apply to the data collection as explained in Section 2.2.4 about programme indicators.

The explanatory sheets include a number of suggestions concerning possible data sources: FADN (Farm Accountancy Data Network), Eurostat's Farm Structural Survey, national statistics, etc).

Such general statistics are in particular useful as context indicators, because they do not only concern the beneficiaries of the programme, but an entire region, population or sector. The suggestions concerning context indicators in explanatory sheets are entirely informative; the responsible authorities or the independent evaluator must define the context information that will be best suited for comparing/adjusting the programme indicators. The above document about common monitoring indicators also includes a limited number of context indicators.

Because the evaluation is effect-wise (opposed to measure-wise) one indicator may sometimes help answer two or more questions, even though the quantification of such an indicator will not necessarily be based on exactly the same population (*e.g. the indicator "Reduction of agricultural inputs per hectare thanks to agreement (%)" is used in both Questions VI.1.B and VI.2.A although not necessarily calculated from the same zones/fields*).

3.3. ANALYSIS

3.3.1. Analysis of programme effects

The analysis of programme effects involves answering the evaluation questions (common and programme specific) as well as other normal evaluation matters such as effectiveness, efficiency, identification of good practice, etc to the extent that this has not yet been covered by the common questions.

The analysis for the common cross-cutting questions will, for example, build on:

- The replies to the chapter specific common questions:
 - particularly the indicators from across the different chapters concerning income, employment, environment, etc

- that some chapters contain no common indicator for certain effects (*e.g., no common environment indicator in Chapter III, training*) does not suggest that this chapter don't contribute to the cross-cutting question about this particular effect (only that a common indicator was not indispensable)
- Systematic consideration of supplier and multiplier effects, negative effects, etc (complementing any such elements from the chapter specific questions).

The evaluation strategy prescribes no particular methodology except that the general quality requirement of Regulation 1750/99 must be complied with as further explained in the Guidelines.

3.3.2. Analysis of the context and of exogenous factors

A systematic analysis of the influence from the contextual situation on the programme should be undertaken as part of the evaluation at national or regional level, as no common questions have been formulated for this issue. This involves various aspects of the setting in which the programme is implemented such as the policy context and the institutional context (including other supporting/regulatory measures and organisations apart from those directly involved in the programme) and other elements of the socio-economic context (*e.g., the agricultural structure within certain sectors or the wider rural context at regional or sub-regional situation*).

The analysis of the overall context of the programme and how it evolves puts the programme into perspective. The context analysis may for example show that a programme, albeit successful in relation to its own objectives, is no longer justified, or on the contrary, that the aid is still justified, even if the programme indicator demonstrates limited progress compared to the objectives. Such information can lead to recommendations at the *mid-term* stage about fortifying or stopping certain measures.

The context analysis can also extract lessons ('best practices') about where/when to apply a certain measure and where/when not to do so.

The context, in the form of exogenous factors, is important for identifying net effects because the programme effects can be entirely obscured by the effects of strong exogenous factors (*e.g., market price, cyclic variations of the economy, weather*) or be very difficult to identify because of modest levels of funding. Comparison to non-assisted peers (*i.e. comparing to the counterfactual situation*) can eliminate the effect of exogenous factors. **Box 3.3** lists the types of situations that may occur regarding the evolution of the beneficiaries compared to that of the context (the example in the box assumes that the context can be used as comparison; sometimes it may be necessary to use other types of comparison groups as discussed in Section 3.1.2).

The explanatory sheets offer preliminary ideas about certain exogenous factors for each individual question, but the evaluation should also comprise an overall consideration of the influences from this type of factors on the programme as a whole.

Box 3.3. Net effect in the presence of significant exogenous factors - typology of situations

CASES OF POSITIVE NET EFFECT

Beneficiary's situation:		Context ^{*)} (sector, region)...	
		improves	deteriorates
Improves...	more than the context ^{*)}	<i>net effect positive (situation B) b %</i>	<i>net effect positive (situation A) a %</i>
	less than the context ^{*)}	<i>no positive net effect (situation E)</i>	
Deteriorates...	less than the context ^{*)}		<i>net effect positive (situation C) c %</i>
	more than the context ^{*)}	<i>no positive effect (situation D)</i>	<i>no positive effect (situation F)</i>

^{*)} Or other comparison groups if the requirements for using the context as comparison group are not fulfilled, c.f., Section 3.1.2

There are three situations with positive net effect:

- Situation A: performance increases for beneficiaries while it decreases for comparison group
- Situation B: performance increases for both groups, but more for beneficiaries than for comparison group
- Situation C: performance decreases for both groups, but less for beneficiaries than for comparison group

There are three situations without positive net effects:

- Situation D: performance decreases for beneficiaries while it increases for comparison group
- Situation E: performance increases both groups, but less for beneficiaries than for comparison group
- Situation F: performance decreases for both groups, but more for beneficiaries than for comparison group

Conclusions:

- The aid can be assumed to have a positive net effect in situations A, B, C: $(a+b+c) \%$
- Of the cases where the beneficiaries have improved their performance (situations A, B, E), only situations A and B indicate positive net effects: $a+b\%$
- In a period of general progress (reference situation improving), the beneficiaries in situation B surpass the trend in their region and period: $b\%$
- In period of general recess (reference situation decreases):
 - The beneficiaries in situation A have improved their individual performance despite of the recession in the region/sector and period: $a\%$
 - the decrease in performance has been moderated compared to general decline in the region/sector for the beneficiaries in situation C: $c\%$