OECD Regional Development Working Papers 2016/02

Using Outcome Indicators to Improve Policies: Methods, Design Strategies and Implementation

Abel Schumann

https://dx.doi.org/10.1787/5jm5cgr8j532-en

OECD

OECD REGIONAL DEVELOPMENT WORKING PAPERS

This series is designed to make available to a wider readership selected studies on regional development issues prepared for use within the OECD. Authorship is usually collective, but principal authors are named. The papers are generally available only in their original language English or French with a summary in the other if available.

OECD Working Papers should not be reported as representing the official views of the OECD or of its member countries. The opinions expressed and arguments employed are those of the author(s).

Working Papers describe preliminary results or research in progress by the author(s) and are published to stimulate discussion on a broad range of issues on which the OECD works. Comments on Working Papers are welcomed, and may be sent to either gov.contact@oecd.org or the Public Governance and Territorial Development Directorate, OECD, 2 rue André-Pascal, 75775 Paris Cedex 16, France.

Authorised for publication by Rolf Alter, Director, Public Governance and Territorial Development Directorate, OECD.

OECD Regional Development Working Papers are published on
http://www.oecd.org/gov/regional/workingpapers

Applications for permission to reproduce or translate all or part of this material should be made to: OECD Publishing, rights@oecd.org or by fax 33 1 45 24 99 30.

© OECD 2016

USING OUTCOME INDICATORS TO IMPROVE POLICIES
METHODS, DESIGN STRATEGIES AND IMPLEMENTATION

By Abel Schumann[1]

Abstract:

This paper discusses the use of outcome indicators for policy monitoring. In addition to providing general recommendations on their design and implementation, it makes two contributions to the existing literature. First, it shows the importance of distinguishing outcome indicators from other types of indicators and demonstrates the need to develop clear policy objectives as a prerequisite for meaningful outcome indicators. Second, it analyses the use of outcome indicators in specific settings; on the sub-national level, in multi-level governance frameworks and in the context of EU cohesion policy. The paper argues that outcome indicators are better used in relational contracts than in transactional contracts between governments. Furthermore, it highlights the need to harmonise different monitoring frameworks within an administration.

JEL classification code: Y80

Key words: indicators, monitoring & evaluation, performance measurement, programme planning.

ACKNOWLEDGEMENTS

This working paper originated from a study conducted by the OECD and the Portuguese government aimed at assessing a set of indicators developed by Portugal. The indicators were developed in order to meet monitoring requirements of the European Commission for projects supported under the EU cohesion policy. The working paper provides a generalised account of the findings of the study. It contains insights developed at a workshop in Lisbon that was organised jointly by the OECD and Portugal's cohesion and development agency and held in October 2014. It was written under the guidance of Joaquim Oliveira Martins and benefitted from comments by Rüdiger Ahrend, Eva Beuselinck, Luiz de Mello, Cornelia Hoffmann, Karen Maguire, Adam Ostry, William Tompson, Wouter van Dooren and participants at a brownbag seminar of the Regional Development Policy Division at the OECD.

[1] Corresponding author: Abel Schumann, abel.schumann@oecd.org.

TABLE OF CONTENTS

INTRODUCTION 5
Focus on results 5
The motivation behind policy indicators 5
TYPES OF INDICATORS AND POLICY OBJECTIVES 6
Distinguishing outcome from output indicators 8
Structuring policy objectives clarifies their importance and their purposes 10
CONSIDERATIONS FOR DESIGNING INDICATORS 12
Avoiding the use of output measures for outcome indicators 12
Matching the indicator to the objective and ensure it is responsive to policy 12
Keeping indicators consistent over time 12
Using data that becomes available without long time lags 13
Normalising indicators with appropriate denominators 13
Minimising the cost of data collection 14
Designing scoreboard indicators 15
Choosing between composite and unitary indicators for scoreboards 15
CONSIDERATIONS FOR USING INDICATORS 16
Using indicators to monitor policies and adjust course when necessary 16
Focusing on objectives and using indicators as tools to achieve them 17
Choosing appropriate benchmark values for indicators 17
Putting indicators into context by using qualitative information 18
Indicators in the presence of policy complementarities 19
Learning and capacity building 19
Making sure incentives based on indicators are aligned with objectives 19
Background information is necessary to understand an indicator 20
Indicators can be tools to inform the public and ensure transparency 20
INDICATORS AS STEERING TOOLS ON THE SUB-NATIONAL LEVEL 21
Sub-national indicators can connect regional policies to national goals 21
Using sub-national indicators to monitor regionally differing effects of national policies 21
Using indicators to achieve better policy coordination 22
Scoreboard indicators on the sub-national level 22
INDICATORS IN INTER-GOVERNMENTAL CONTRACTS 23
Types of inter-governmental contracts 24
Outcome indicators in relational contracts 24
Outcome indicators in transactional contracts 26
INDICATORS IN THE CONTEXT OF EU COHESION POLICY 27
Aligning different monitoring frameworks 27
CONCLUSION 29
BIBLIOGRAPHY 30

Figures

Figure 1. Types of Indicators 7
Figure 2. Distinguishing Outcome and Output Indicators 9
Figure 3. A Schematic Hierarchy of Policy Objectives 10
Figure 4. Determinants of policy outcomes 25

INTRODUCTION

Focus on results

In recent years there has been a trend towards defining clear objectives of policies and assessing their performance. Instead of implementing policies and hoping that they will have the desired consequences, governments make systematic attempts to measure their effects and adjust them for greater effectiveness. Within this broader trend towards evidence-based policy making indicators play an important role as a tool for policy monitoring.

For the purpose of this paper, indicators are defined as quantitative or categorical measures that provide information on conditions and developments that are relevant for the policy making process. Depending on their function, different indicators have different uses. In their most basic function, they provide information on the baseline situation in a policy field. In more elaborate settings, they serve to monitor the resources spent on a policy, the efficiency a policy is implemented with or the effectiveness of a policy in achieving its objectives. While all good indicators share some characteristics, other properties of indicators depend on their uses.

This working paper focuses on outcome indicators – the most important type of indicator for policy monitoring. It develops a framework how differentiate outcome indicators from other indicators, how to design them and how to use them in different contexts. First, it provides an overview of the main types of indicators and shows how their interpretation differs from each other. One of the defining characteristics of outcome indicators that are discussed in this context is their close connection to policy objectives. Second, the paper provides recommendations on how to design and use outcome indicators. Although an extensive literature on the topic exists, the paper provides a novel angle by highlighting the specificities of outcome indicators compared to other types of indicators. Third, the use of outcome indicators in specific settings is discussed. In particular, the paper analyses the use of indicators as a tool for sub-national governments, for contracts between governments in a multi-level governance framework and within EU cohesion policy. These particular roles of outcome indicators have not been widely discussed in the literature, yet.

The motivation behind policy indicators

It is never possible in advance to say with certainty what the effects of a policy will be. Ideas that sound good in theory can disappoint in practice, whereas others turn out to be unexpectedly successful. Even policies that have been tried in the past can have different effects than before when seemingly unrelated circumstances have changed in the meantime. In general, modern policies have such complex effects and are so strongly interdependent that a priori their exact consequences can never be predicted with certainty. Therefore, policies have to be monitored and evaluated to determine what their effects are. Without systematic monitoring and evaluation, it is almost impossible to separate policies that work well from those that do not.

Indicators play a central role in the monitoring process by generating regular and objective feedback about progress towards policy objectives. They are quantitative representations of the conditions in a policy field. They can be used as a tool to examine the effects of policies and they provide crucial information for policy makers to judge the effectiveness of policies and to make adjustments where required. Compared to many other feedback mechanisms, well-designed indicators have the advantage that

they provide easily comprehensible information. Thereby, they can form a factual basis upon which informed political decisions can be made.[2]

Despite the obvious importance of indicators, they are not ends in themselves but rather instruments to ensure that policies achieve the desired outcomes. The focus of a policy should always be on the policy objective and never on an indicator value itself. Concentrating on improving an indicator value instead of achieving the objective runs the risk that policies are tailored to do well on the indicator but perform badly in terms of achieving their broader objective. This risk becomes even greater if financial or other performance incentives are based on indicators. Therefore, when using indicators it should regularly be scrutinised whether policies are still aligned with their actual objective.[3]

TYPES OF INDICATORS AND POLICY OBJECTIVES

Types of indicators: What is measured and for which purpose?

Indicators can be classified into three general categories according to what is measured; input indicators, output indicators, and outcome indicators. Even though this distinction is well-established and widely mentioned in the related literature, it is worth summarising it briefly because it is a defining characteristic of an indicator. Input, output and outcome indicators can all be part of a framework to monitor the effectiveness of policies, but it is important that each type of indicator is used to monitor only the aspects of a policy it is designed to monitor. In particular, it is essential to avoid using output indicators in order to measure outcomes. Figure 1 provides an overview of the three types of indicators.

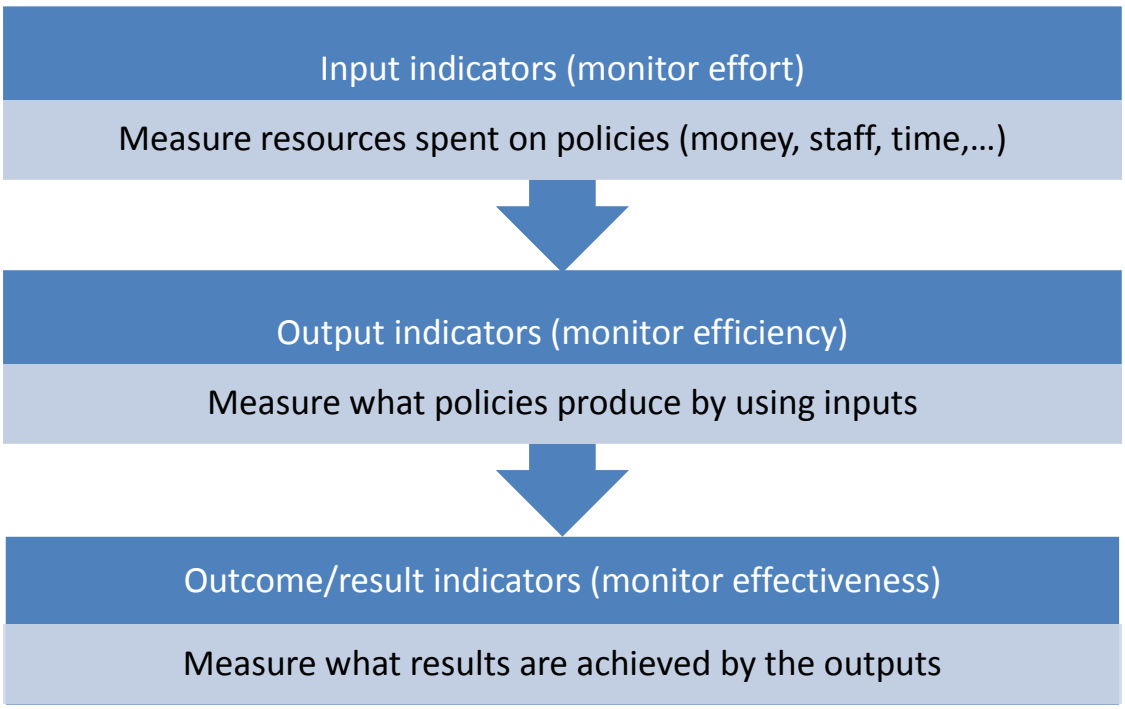
Input indicator

Input indicators are used to measure the amount of resources that are allocated to a policy. Thus, they are measures of effort. A typical input indicator is the funds that are spent on a certain policy or the number of people working on a project. Input indicators can provide a measure of the effort that is devoted to pursuing a policy. They do not give any information whether the resources are efficiently spent or whether a policy is effective in achieving an objective. The role of input indicators in monitoring is limited to providing information on the intensity with which a policy is pursued. Therefore, they should usually be employed in combination with output indicators and outcome indicators. It is never possible to monitor policies comprehensively using input indicators alone.

[2] See OECD (2009) and Van Dooren, Bouckaert and Halligan (2015) for a discussion on the uses of performance indicators.

[3] For a detailed discussion of the arguments presented in this section, see McCann (forthcoming).

Figure 1. Types of indicators



Source: Authors' elaboration.

Output indicator

Output indicators monitor how efficiently policies are executed. The information they provide can help to improve the implementation of policies. Output indicators measure quantities that are produced by a policy in order to achieve its objectives, but not progress towards the objectives itself. The importance of output indicators depends on the policy in question. They are particularly important for policies that are implemented with varying efficiencies, but less important if policies are straightforward to implement. Some specific policies (such as tax policies) do not create outputs and no output indicators exist for them.

Outputs are means to achieve a policy objective, but no ends in themselves. They are produced because policy makers expect them to contribute to desirable outcomes (see Figure 2). Typical output indicators might show the number of motorway kilometres built, the number of people trained to fulfil a task, or the area for which environmental protection plans have been developed. Output indicators do not provide any information whether the outputs of a policy are effective in achieving the desired outcomes of a policy, i.e. if the outputs do what they are supposed to do. Consequently, the primary purpose of output indicators is to monitor if a policy is efficient in producing outputs. They cannot monitor the effectiveness with which the policy addresses its objective.

Outcome/result indicator[4]

Outcome indicators are used to monitor the effectiveness of policies in achieving their objectives. They help to understand whether policies are well-designed in view of their objectives. Outcomes are the underlying motivation behind policies, but in most cases they can only be affected through the production of outputs. Typical outcome indicators might be the reduction in travel time between cities, the increase in customer satisfaction with a service provision, or the number of species no longer threatened by extinction. All these outcomes have in common that they cannot be changed directly. There is no law, which can decree that a species is no longer threatened by extinction. In order to affect this outcome, a policy has to produce an output that influences the outcome in the desired way.

An outcome indicator always has a normative component in the sense that (within a reasonable range) a movement in one direction is considered a positive development and a movement in the other direction is considered a negative development. Some publications distinguish between outcome indicators that measure inherently desirable outcomes and those that are desirable only because they contribute to a higher level outcome.[5] This paper does not make the distinction and defines outcomes simply as the results that are supposed to be achieved by the outputs of a policy.

Distinguishing outcome from output indicators

Distinguishing outcome from output indicators and using them according to their purpose is a crucial aspect when using indicators to monitor policies. Generally, outputs refer to something that a policy produces directly with the inputs that are provided. In contrast, outcomes are something that should be achieved by the outputs that are produced.

Distinguishing outcome and output indicators is easiest if it is exactly specified *what* a policy produces and *why* it produces it. In this case an indicator that measures the *what* is an output indicator and an indicator that measures the *why* is an outcome indicator. Although this rule is generally valid, it is only useful if the *what* and the *why* are properly specified. The policy objective must clearly distinguish between what a policy produces and what motivates it. If this is not done, there is still a risk that output and outcome measures are confused with each other.

Usually, there is certainty about what outputs are produced by a policy, but it is generally not certain if the outputs cause a certain outcome. For example, a policy that aims to build new schools is virtually guaranteed to produce new school buildings as outputs. However, it is not clear if the new schools improve the learning outcomes of students. Outcome indicators always measure something that is not perfectly assured to be achieved by a policy (thus the need for them). There might or might not be better learning outcomes as a result of school construction. In contrast, if there is any uncertainty related to outputs, it is usually only regarding the quantity and quality of outputs produced. Even though it is clear that a policy to build schools will produce new school buildings, it might not be clear how many will be built nor of what quality they will be.

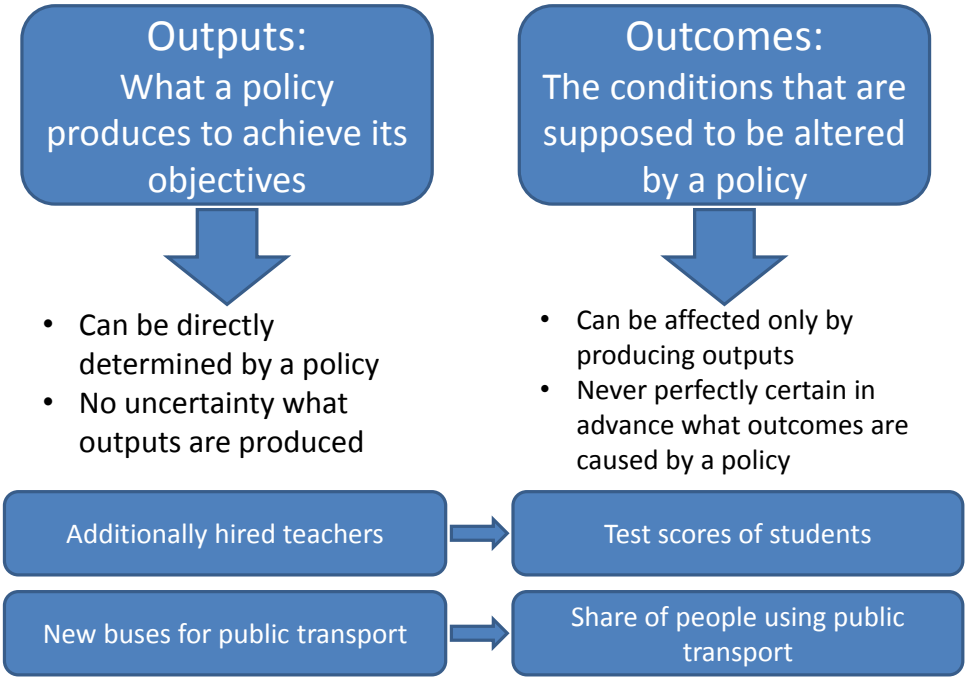
---

[4] The terms “outcome indicator” and “result indicator” usually have an identical meaning. “Outcome indicator” is more commonly used in the literature on indicators, whereas the European Commission speaks of “result indicators”. In keeping in line with most of the existing literature on indicators, this paper refers to them as “outcome indicators”.

[5] Unfortunately, there is no generally accepted terminology how to refer to them. In some cases, outcome indicators that do not measure an inherently desirable (i.e. normative) outcome are called intermediate outcome indicators. In these cases, outcome indicator without the suffix “intermediate” usually refers to outcome indicators that have an explicit normative component. In other cases, indicators that refer to outcomes with direct normative connotations are called impact indicators and indicators that refer to outcomes without explicit normative components (i.e. those outcomes that only serve to achieve other outcomes) are called outcome indicators.

Outcomes are often physically immaterial (such as the amount learned by a student, the quality of a product or a service, the health of a group of people, the customer satisfaction, or the performance of an organisation. In contrast, outputs are more commonly material things that exist physically, such as the length of new roads, the number of people hired or the number of facilities build. This distinction can prove helpful in distinguishing outcome and output indicators. However, it is not a strict rule. There are also outcomes that are material and outputs that are immaterial.[6] For example, many services provided by the state are outputs even though they are immaterial. Therefore, it should not be the only criterion to distinguish the two types of indicators.

Figure 2. Distinguishing outcome and output indicators



Source: Authors' elaboration.

Policy objectives are a prerequisite for meaningful outcome indicators

Indicators that monitor policies are supposed to provide information about their effectiveness. In order to judge the effectiveness policies, criteria against which to judge them are needed. This is the role of policy objectives. Indicators are supposed to monitor the effectiveness of policies in achieving their objectives, i.e. in doing the things they were designed to do.

As outcome indicators measure the effectiveness of a policy in achieving its objectives, it is not possible to develop meaningful outcome indicators without having previously specified policy objectives. Usually, every policy has an objective but it is not always made explicit. Even if explicit policy objectives exist, they are not always clear about the outcome that they aim to influence. The more explicit policy objectives are in this respect, the easier it is to develop appropriate outcome indicators. If objectives are ambiguous about the outcomes for which they aim, it makes it difficult to develop appropriate indicators

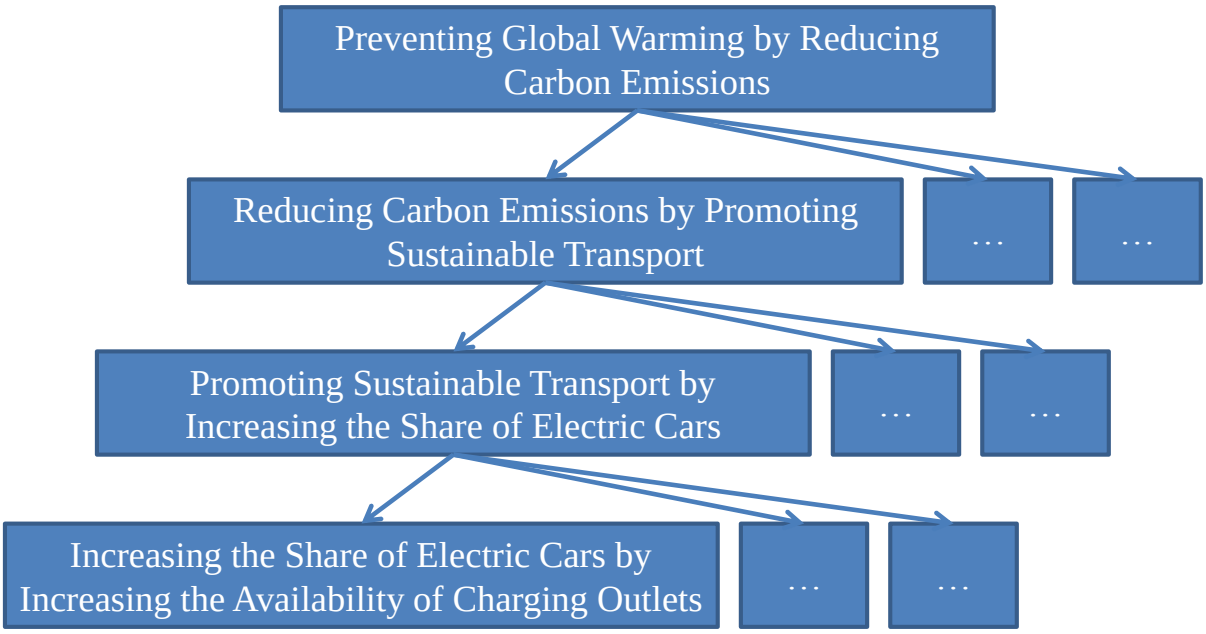
[6] For example, strategic plans can be outputs of policies, but are immaterial. Conversely, threatened species that have been saved from extinction are material things but are generally outcomes of policies.

because it is unclear which outcome should be measured. In contrast, if the outcome of an objective is clearly stated and quantifiable, the corresponding outcome indicator often follows immediately from the objective. Even if the outcome of an objective is not directly quantifiable, a clear definition of the outcome nevertheless helps to identify suitable proxies for the outcome that can be used as indicators.[7]

Structuring policy objectives clarifies their importance and their purposes

Not all policy objectives have the same importance. Typically, they can be ordered in a hierarchical fashion where one objective contributes to achieving another objective at a higher level. Being aware about the hierarchical order of objectives can facilitate the development of appropriate indicators. In order to make this point clear, it is useful to consider an example. Figure 3 shows an example of several policy objectives that are ordered hierarchically. The overarching objective is to reduce carbon emissions to prevent global warming. There are many different policies that can contribute to achieving this objective. Among them is the strategy to promote sustainable transport. As shown on the second level, this strategy can also be phrased as a policy objective. Again, there are many different policies that contribute to the promotion of sustainable transport. One of them is to increase the share of electric cars, the objective at the third level. As in the previous cases, many policies can contribute to achieving the objective. Among them are policies aiming at increasing the availability of charging outlets, as specified in the fourth objective. Generally, the lower an objective in the hierarchy, the more easily can it be affected by a single policy.

Figure 3. A schematic hierarchy of policy objectives



Source: Authors' elaboration.

The graph above is just a schematic representation. In practice, the relation between different policy objectives is usually more complex. Just as a higher level objective requires the contribution of several lower level policies to be achieved, often a lower level policy contributes to achieving several higher level objectives. Increasing the share of electric cars does not only contribute to promoting sustainable transport and reducing carbon emissions, it also increases the resilience of the economy by lowering its dependence on imports of oil and improves the air quality in urban areas by reducing pollutants from exhaust fumes.

[7] Proxies are measurable variables that are so closely related to non-measurable or non-quantifiable variables that they can be used in place of the non-measurable variable for analytical purposes.

Furthermore, policy objectives on the same horizontal level are often complementarities – a good performance in one dimension improves the performance in another dimension and vice versa. In other cases, policy objectives are substitutive – a good performance in one dimension makes it more difficult to achieve a good performance in another dimension.

Each of the four policy objectives shown in Figure 3 can be monitored by an outcome indicator. However, the figure makes clear that the four corresponding indicators would not be directly comparable to each other. The further up an indicator is located on the hierarchy, the more closely it monitors progress in achieving the strategic goals behind a set of policies. The further down an indicator is located in the hierarchy, the more closely it monitors the effectiveness of a specific individual policy. Indicators that fall in between the two ends typically monitor how effective a set of policies is in contributing to an intermediate objective.

An indicator should always measure an outcome that is on the same hierarchical level as the policy objective it is supposed to monitor. In other words, if an indicator is supposed to monitor the objective "Promoting Sustainable Transport by Increasing the Share of Electric Cars", neither the number of public charging outlets nor the amount of carbon emitted from the transport sector is a suitable indicator. The first is at a level below the objective and measures only one of many aspects related to achieving the objective. The second is at a level above the objective and is influenced by a variety of other policies that are not directly related to the objective. Neither variable is a suitable indicator for the particular objective, but would work well as an indicator for a different policy objective.

Indicators at all levels of the hierarchy described above can contribute to inform policy. When indicators from different levels of the hierarchy are used simultaneously it can help to obtain a better understanding how different policies contribute to achieve outcomes at different levels. However, it is always important to keep in mind what objective an indicator is monitoring. In particular, simplistic interpretations that do not differentiate between indicators (such as "x indicators are moving in a positive direction and y indicators are moving in a negative direction, therefore…") should be avoided because they are misleading. Instead, the vertical and horizontal relations between the different outcomes should be carefully considered when interpreting the indicators.

A system of indicators that takes the relations between objectives into account can contribute to a better understanding of the reasons behind policy successes and failures. Whereas individually indicators provide feedback on specific policies, a system of indicators that matches the relations of objectives can provide feedback on entire policy strategies. In particular, it can help to show which policies are successful in contributing to strategic objectives and which are bottlenecks that prevent progress.

Going beyond a system of indicators, programme logic models can be an important step towards a better understanding of performance indicators. They are descriptions of how policies affect outcomes that can include more details than representations entirely based on indicators. Due to this, logic models can help to obtain a better understanding of the reasons why policies work or do not work. If outcome indicators show that a policy is not working, logic models can inform decision makers how to change course effectively. Furthermore, they can provide the basis for a dialogue among involved actors on how to achieve objectives effectively. [8]

[8] See for instance the report of the Kellog Foundation on logic models. Kellogg Foundation, W. K. (2006). Using logic models to bring together planning, evaluation, and action: Logic model development guide. https://www.wkkf.org/resource-directory/resource/2006/02/wk-kellogg-foundation-logic-model-development-guide.

CONSIDERATIONS FOR DESIGNING INDICATORS

Indicators[9] are always tailored to the outcome of the policy that they are supposed to monitor. Therefore, their quality can only be judged in the context of specific policy objectives. Nevertheless, there are some general considerations that should be taken into account when devising indicators.

Avoiding the use of output measures for outcome indicators

The previous section has shown the difference between output and outcome indicators and has argued that both types of indicators have their justification. However, their use needs to match their respective purpose. Outcomes should always be measured by outcome indicators and outputs by output indicators. Although this sounds obvious, one of the most common mistakes in devising indicators is the use of an output measure for an outcome indicator. Therefore, the careful verification that a proposed outcome indicator actually represents an outcome measure and not just an output is an essential part in the development process of an indicator.

Matching the indicator to the objective and ensure it is responsive to policy

An ideal indicator captures the essence of the policy objective while minimising the influence of confounding factors. In practical terms, this implies that an indicator should be as little as possible influenced by factors that lie outside of the policy objective, but should not ignore any aspects of the objective. Indicators that are too narrow leave out important information, whereas indicators that are too broad are unreliable because they are influenced by too many factors that are not directly relevant for the policy objective.

Indicators need to be well-aligned with objectives in the sense that the outcome that is measured by the indicator is the outcome that a policy is aiming at. In other words, the indicator should measure an outcome that can be reasonably expected to be affected by a well-designed policy. As most policy objectives are complex and often only partially quantifiable, it is normal that there are discrepancies between the outcomes of policy objectives and the outcomes measured by indicators. However, if the differences become too large, there is a risk that an indicator measures an outcome that is not responsive to the policy in question. Such an indicator would be ineffectual for monitoring and should therefore be discarded. In order to find the appropriate indicator for an objective, the hierarchical maps of policy objectives described earlier and programme logic models are helpful tools. They can clarify how objectives and policies are related to outcomes and indicators.

Keeping indicators consistent over time

Monitoring developments over time is a central function of indicators. Looking at an indicator for a single time period can provide valuable information, but the full potential of indicators is realised only by analysing how they develop over time.

In order to ensure that consecutive readings of an indicator are comparable, changes in the definition of indicators should be avoided whenever possible. The need to change the definition of indicators can be reduced by considering two aspects when devising them; Indicators should be based on data that is available in regular time intervals and they need to refer to an outcome that remains relevant for the foreseeable future.

[9] Unless otherwise mentioned, in the following the term “indicators” refers to outcome indicators. This notwithstanding, most of the discussion is also applicable to any other type of indicator.

Some data is routinely collected by statistical agencies or other public organisations. For such data, regularity is no concern as the data usually becomes available at yearly or higher frequencies. However, other potential data sources do not always become available regularly. Some potential outcomes, such as customer satisfaction, are best measured by surveys. Many surveys are one-off data collections that were not initially designed to be repeated on a regular basis. When considering such data sources for an indicator it should be taken into account whether it is possible to collect the data frequently enough to be useful as an indicator.

Even if data is available regularly, it is not assured that the data will remain relevant in the future. Some outcomes are specific to particular events. Examples of such outcomes are the length of cleaned beaches after the sinking of an oil tanker or the number of re-trained workers that became unemployed in a large bankruptcy. Indicators should refer to such outcomes only if the policy objective is specific to the particular event. In other words, if the policy objective is to contain the environmental damages from the sinking of the oil tanker, the share of cleaned beaches might be a useful indicator. However, if the policy objective is to reduce maritime pollution, the share of cleaned beaches is most likely not a good indicator. Even if it is an important outcome immediately after a disaster, the share of cleaned beaches would become less and less important for the overall objective as time passes and thereby less and less relevant as an indicator.

Using data that becomes available without long time lags

A central function of indicators is to provide continuous and timely feedback to policy makers. To be able to accomplish this task, indicators need to be based on data that becomes available quickly and at a sufficiently high frequency. Often, data becomes available with a considerable time lag. It is not unusual that it takes one year or more for data to be published and even then it is often subject to later revisions, which can change its interpretation substantially. Depending on the time frame of the policy objectives, such time lags can severely affect the potential of an indicator to inform policy decisions. In particular, policies that have short-run objectives or need frequent adjustment require timely feedback because decisions have to be taken quickly. Indicators that are supposed to provide guidance on such objectives need to become available without much delay.

When judging the quality of potential indicators, the timeliness with which they become available should always be considered. The quicker information becomes available, the faster policy makers can react to it. While in general it is always preferable to get data as soon as possible, there is no universal rule that specifies which time lag is acceptable and which is unacceptable. Timeliness is one of several quality criteria of an indicator. Its importance depends on the policy field that an indicator monitors and can vary depending on how frequently policies need to be adjusted.

Normalising indicators with appropriate denominators

Indicators are supposed to provide policy makers and the public with easily accessible and meaningful overviews of the conditions in different policy fields. Therefore, indicators should convey as much information as possible in a single number. Furthermore, as little additional information as possible should be required for their interpretation.

This implies that indicators should be normalised with suitable denominators. Normalisation in this context refers to the division of the actual outcome variable by another variable. Usually, the aim is to express the indicator per unit of another variable or as share of a larger category. Typical examples of such normalised indicators are “carbon emissions per unit of GDP”, “the number of patent applications per inhabitant”, “the share of students obtaining at least a high school degree”, and “the share of renewables in

total electricity consumption". All these indicators have in common that the outcome could be expressed in non-normalised terms, but for various reasons it is desirable to present them as mentioned above.

The most obvious reason is that it is often easier to understand the importance of a number if it is expressed in normalised form. Most people will find it easier to judge the importance of the percentage of a country's electricity consumption that came from renewable sources than the number of terra watt hours (TWh) of renewable energy that a country consumed. Both measures use the same outcome (i.e. electricity consumption from renewables), but in the first example, the outcome is divided by the total electricity consumption, whereas in the second example it is presented without any further modification.

Second, normalised indicators are better suited for comparisons. Comparing a country's high school graduation rate of X per cent to another country's graduation rate of Y per cent makes sense, but comparing the total number of the countries' high school graduates to each other is usually meaningless due to the different sizes of the countries' student population.

Third, normalised indicators are less strongly influenced by contextual developments that are not related to the policy that is monitored by the indicator. Often, fluctuations in other factors affect the outcome that is supposed to be monitored. Dividing the outcome by a variable that measures the fluctuating factor removes its influence on the indicator. Carbon emissions for example tend to fluctuate with the business cycle. Due to increased economic activity they rise during booms and decrease during recessions. Using total carbon emissions as an indicator for carbon efficiency of the economy would therefore give the false impression that a country is doing badly in this respect during booms and well during recessions. Dividing carbon emissions by GDP (i.e. using carbon emissions per unit of GDP) would provide a measure of carbon efficiency that is less sensitive to overall fluctuations in GDP.

When using denominators that fluctuate independently from the nominator, carefully attention has to be paid to what causes a change in an indicator. It could be the result from changes in the nominator or changes in the denominator. For example, an increase in the share of businesses that export could be the result of more businesses exporting or it could be the result of non-exporting businesses going bankrupt. Whereas both developments would result in an increase in the indicator, only the first situation would be desirable. If fluctuations in the denominator frequently lead to an improved reading of an indicator despite actually negative developments, it might be preferential to use a denominator that is more stable or to use absolute values instead of fractions.

Although it is advisable in most cases to normalise an indicator, there are exceptions to the rule. In some cases, it is the absolute value that is policy relevant. For example, a community might rely on water extraction from a river for its water supply and there is only a certain annual amount of water that can be extracted from the river without harming its ecosystem. In this case, the absolute amount of water extraction is clearly the more important variable for assessing the consequences for the river ecosystem even though water extraction per capita gives a better impression of the efficiency of water usage in the community.

Minimising the cost of data collection

The administrative burden and costs related to collecting and processing relevant data can be significant. They should always be kept as low as possible and be justified by the desired use of the indicator. Whenever possible without sacrificing the quality of an indicator, existing data sources should be used for it. If an indicator requires new data collection, it needs to be weighed up whether the gain in quality due to the new data justifies the costs of collecting it. If the information gain from a new indicator is small or the collection of data very costly, it might even be reasonable to forego the use of an indicator completely.

In this context, it is important to consider not only the costs for the public administration, but also the costs imposed on the private sector and civil society. Data collection and other information provision requirements create significant costs for businesses. Whenever possible, governments should aim at reducing this burden in order to foster private sector activity.

Designing scoreboard indicators

Scoreboard indicators are the highest level of outcome indicators. They have a special role in a system of indicators. Primarily, they serve communication purposes and are supposed to provide general overviews of the conditions in entire policy fields. In contrast to lower level objectives, they do not aim at documenting the consequences of individual policy measures. Instead, their function is to showcase the effects of broad government strategies.

Because of the generality of scoreboard indicators, it is difficult to derive concrete policy proposals from them. Scoreboard indicators can be useful in demonstrating *if* a general strategy is working, but they cannot provide any information on *why* it is working or not. In order to analyse this question, indicators that are more closely related to individual policies are more appropriate. Despite their limitations, scoreboard indicators have several advantages. They can serve as a yardstick to judge the performance of the government. They can also be a tool for the government to highlight the importance of its priority policy fields to the public. If scoreboard indicators are widely accepted throughout different governmental agencies, they can also serve as a focal point to concentrate efforts and align different programmes around them.

Like all other outcome indicators, scoreboard indicators need to be based on policy objectives. Policy objectives clarify the priorities of the government internally and externally and justify the selection of a particular indicator as a quantitative measure that corresponds to the objective. In the case of scoreboard indicators, the setting of objectives is especially important because scoreboard indicators send a strong signal about the priorities of the government and should therefore correspond to its actual priorities.

Depending on the complexity of the policy field, some objectives for scoreboard indicators require much more detail than objectives for other types of indicators. Some policy fields cover a wide range of specific objectives and related outcomes that can contribute to it. Based on technical criteria alone, it is not possible to judge the importance of the different contributing objectives. Before an appropriate scoreboard indicator can be defined, political decisions regarding the importance of the different specific objectives have to be made and should be spelled out explicitly.

Choosing between composite and unitary indicators for scoreboards

Scoreboard indicators differ from other outcome indicators primarily in their level of detail. As they are expected to summarise entire policy fields, it is often difficult to find a single measure that contains all the relevant information to represent the policy field adequately. Composite indicators that combine several measures into one indicator avoid this problem. However, they also have significant downsides and should be used only carefully.

Some policy fields can easily be represented by unitary indicators. For example, the strategic objective to avoid climate change is reasonably well represented by an indicator that measures CO2 emissions.[10] However, other strategic policy fields are more difficult to represent by a single indicator. As mentioned above, there is no single unitary indicator that encompasses all the different dimensions of the

[10] This indicator would ignore the impact of other greenhouse gases, but the simplification could justified be given the importance of CO2 for climate change.

objective "Promoting Social Inclusion and Fighting Poverty". In this case, composite indicators could be used to summarise different outcomes into a single indicator.

Despite their advantages, composite indicators can easily be misleading and should therefore be used only after careful consideration. Due to the technical complexities of constructing composite indicators, several problems might occur. First, composite indicators can be misinterpreted if the details of their construction are not taken into account or are not fully understood by the public. Second, composite indicators can be very sensitive to the way they are constructed. Whether they move in a positive or negative direction can be primarily due to their construction and not due to the fundamental developments in the policy fields they monitor. Third, the sensitivity with regard to different specifications can be exploited for political purposes. It is possible to construct seemingly sensible composite indicators that tend to show a politically desired outcome. Instead of fostering transparency, such composite indicators are intentionally misleading and reduce transparency.

The problems mentioned above can be avoided by choosing internationally recognised composite indicators. These tend to use established methodologies and cannot be influenced for political purposes. For example, the PISA score is a composite indicator that aggregates the outcomes of different tests administered as part of the OECD's PISA study. It is widely regarded as a valid and objective measure of learning outcomes of students and could thus be an appropriate scoreboard indicator.

If composite indicators are nevertheless created from scratch, it is important to use adequate statistical methods. Among the concepts that are important for the construction of scoreboard indicators are normalisation, aggregation, weighting and sensitivity analyses. A discussion of them is beyond the scope of this paper. Interested readers are referred to the OECD/EU/JRC Handbook on Constructing Composite Indicators that contains a comprehensive discussion of the methods required to construct composite indicators.[11]

CONSIDERATIONS FOR USING INDICATORS

Using indicators to monitor policies and adjust course when necessary

As stated previously, indicators are an essential tool for monitoring policies and they should be used accordingly. A well-functioning monitoring framework provides feedback that helps to judge its effectiveness and to make course-adjustments when necessary. That said indicators should always be used in combination with other information.

Even though indicators are valuable for judging the effectiveness of a policy, they can never provide a definite answer whether a policy is effective. All outcome indicators are influenced by external factors that are only indirectly related to a policy. It is never possible to know exactly what part of the change in an indicator is due to a policy and what part is due to external factors. Therefore, indicators cannot provide a precise estimate of the consequence of a policy. If an indicator shows a negative development of an outcome, one possible explanation is an ineffective or even counterproductive policy. Another explanation is that the policy works well and the outcome would have been even worse if the policy had not been in place. This shows that indicators always need to be assessed in light of other developments. The information contained in an indicator generates its full value only if it is combined with contextual information. Such information could come from other indicators, but in principle all types of information can be useful to interpret indicators.

[11] See OECD/EU/JRC (2008).

Policy monitoring provides valuable feedback to policy makers, but it cannot give a definitive answer regarding the effectiveness of a policy. In order to do so, different approaches are required that are commonly summarised under the label *policy evaluation*. Policy evaluation is usually based on statistical and econometric techniques that distinguish the impact of a policy from confounding factors, (i.e. can identify the causal effect of a policy). Policy evaluation has typically higher data requirements than policy monitoring and often requires that policies are designed in specific ways that allows them to be evaluated.[12]

Ideally, monitoring and evaluation should complement each other. Monitoring is supposed to provide a continuous feedback that shows strengths and weaknesses within policy fields and enables policy makers to react quickly to them. In contrast, the high analytical requirements of policy evaluation imply that in most cases it is impossible to do it regularly and in short intervals. However, if it exists, a properly conducted policy evaluation provides a higher certainty about the effectiveness of a policy than monitoring by indicators.

Focusing on objectives and using indicators as tools to achieve them

Indicators are tools that can help to ensure that policies achieve their objectives, but they should not be regarded as objectives themselves. Policies should always put priority on achieving an objective, not on achieving a good indicator reading. Policies that focus on indicators instead of objectives cause two problems. First, they might be ineffective or even counterproductive in achieving an objective. Second, they can create the false impression that conditions in a policy field are more favourable than they actually are and make policies wrongly look successful.

Even well designed indicators are rarely perfectly representative of a policy objective. As a consequence, it is often possible to selectively improve an indicator while contributing much less to the actual policy objective. For example, the share of university graduates within a cohort is a reasonable indicator for the objective "Promote Higher Education to Increase the Qualification of the Labour Force". A sound policy that focuses on the objective could aim at improving conditions at universities or at providing financial support for low income students. A misguided policy that focuses only on the indicator could lower requirements for graduation in order to increase the number of graduates.

These problems can partly be mitigated by the design of indicators. The more closely an indicator tracks the outcome that should be achieved, the less likely it is that a policy, which improves the indicator, is ineffective or counterproductive with respect to the actual objective. Nevertheless, when using indicators, it should always be questioned whether a policy primarily serves to achieve a desired outcome or to improve an indicator.

Choosing appropriate benchmark values for indicators

Even well-designed indicators need to be compared against reference values to interpret them. Furthermore, benchmark values can serve as quantitative specifications of objectives. The need for benchmark values can be easily demonstrated with an example. Assume that the innovation activity of the private sector should be measured and R&D investment of businesses in per cent of GDP is used as indicator. Although private R&D investment as share of GDP is a perfectly valid outcome indicator, it is not very informative without being put in context. Very few people have an intuitive understanding whether a value such as 0.69 per cent of GDP is a large or a small share for R&D spending. In order to understand the implications of the indicator, it has to be compared against relevant reference values.

[12] For an in-depth discussion and advice on techniques for policy evaluation, see the guidance documents provided by the European Commission. http://ec.europa.eu/regional_policy/information/evaluations/guidance_en.cfm#1.

There are several methods to develop benchmark values. They are not mutually exclusive and, if possible, should all be considered when deriving benchmark values. In some instances, relevant benchmarks are provided by scientific insights. For example, a benchmark for air pollution with particle matter PM2.5 could be the guideline value for a low risk exposure provided by the World Health Organization (i.e. a daily average not exceeding 25µg/m³ and an annual average not exceeding 10µg/m³). It would be obvious to consider any value below the threshold as satisfactory and any value above the threshold as unsatisfactory.

Benchmarks can also be developed based on historical data in order to compare current performance with past performance. Ideally, benchmarks based on historical data are based on an average of several past observations. This prevents them from being excessively influenced by outliers that were caused by random fluctuations at a single point in time. Furthermore, when the historical data shows a trend, this should be taken into account when developing benchmark values. For example, an indicator that has been growing with 5 per cent per year on average in the past might be expected to continue to grow at this pace without any policy intervention. Any meaningful benchmark value would therefore need to take this regular growth into account.

Comparisons with other countries are another possibility to derive benchmarks. This is especially important if indicators are collected for the first time. In this case, little or no historical data will be available to analyse the development of indicators over time. Possible data sources for international comparisons are the databases of international organisations such as the OECD and the World Bank, of Eurostat and of national statistical offices.

Whenever international data sources are used, particular attention should be paid to the comparability of the data. Often, the definitions behind data series vary from country to country. This can limit the comparability of the data even if it is supposedly the same statistic. For example, national definitions of the unemployment rate differ strongly. Individuals that are considered unemployed in some countries are not counted as unemployed in others. In such cases, it is preferable to use data series with a common definition. If this is not possible, it should be assured that, despite different underlying definitions, the data is comparable across countries.

When deciding with which countries to compare the own performance, two considerations have to be made. On the one hand, well-performing countries should be selected to provide incentives to catch up with the best performers. On the other hand, socio-economic conditions in the comparison countries should not differ too strongly from the own country to allow for meaningful comparisons.

Constructing benchmarks is difficult if no scientific insights regarding appropriate values exist and neither historical data nor data from other countries is available. In this situation, it is appropriate to wait with the formulation of benchmarks and targets until the first data becomes available. If it is imperative to develop benchmark values immediately, available data for similar outcomes and programmes could provide help in establishing them.

Putting indicators into context by using qualitative information

The use of qualitative and contextual information is an essential complement to the use of indicators. Well-designed indicators have the advantage that they provide an objective picture of circumstances. However, they can never provide a complete picture of a situation in a policy field. In addition to quantitative information, a wide range of qualitative information exists that should be systematically collected and analysed. Such qualitative information provides context that can help to better understand the mechanisms through which policies work and how they interact with each other.

Indicators in the presence of policy complementarities

Most policies do not work in isolation. Their effectiveness generally relies on conditions in areas that are not directly related to them and apparently unrelated policies can be crucial for the success or failure of a policy. These so-called complementarities present important opportunities, but also challenges to policy makers. In the best case, policies that address complementary policy fields simultaneously can create virtuous cycles, in which progress in one field leads to progress in another field and vice versa. In the worst case, a well-designed policy can be rendered ineffectual by the lack of a complementary policy.

Ideally, when such complementarities are known to exist, they can be measured by other indicators and policy makers can take them directly into account when making decisions. Often, however, it is not known what complementarities exist. In these cases, it is important not to jump to the conclusion that a policy does not work if an indicator does not show improvement. It might be the case that a complementary policy is lacking. In general, using systems of indicators that represent the relations of different policy objectives to each other can be helpful in identifying complementarities and to address them jointly. Furthermore, the abovementioned use of qualitative and contextual information can contribute to a better understanding of how a policy works. This can help to identify complementary policies that need to be in place in order to ensure the policy in question is effective.

Learning and capacity building

Indicators are supposed to monitor policies so that they can be altered if they are not working. They also provide the opportunity for learning about the effectiveness of policies in a broader context. By providing continuous feedback on the conditions in a policy field, indicators help to gain a better understanding of how policies work and to build a stock of knowledge about characteristics of successful policies. Thereby, they contribute not only to improving the policies that they monitor, but can also improve future policies. In order to facilitate the learning process, the lessons learned from using indicators should be shared within the public administration.

In order to maximise the learning potential from an indicator system, it is desirable to analyse indicators not only within the on-going policy making process, but also ex-post, i.e. after all relevant policy decisions have been made. Such an analysis can help to repeat successes and avoid repeating mistakes.

Making sure incentives based on indicators are aligned with objectives

Due to their quantitative nature, indicators facilitate the setting of performance incentives. Indicators are objective and unambiguous about whether or not a target has been met. Therefore, many incentive schemes rely on indicators to provide target values.

While incentive-based policies can be valuable tools to maximise efforts from stakeholders to achieve objectives, there is a risk that incentives are misaligned whenever the indicator is not a perfect representation of the desired outcome. In particular, it needs to be ensured that incentivised stakeholders do not try to game the system by taking steps that aim to improve the indicator without promoting the actual policy objective. This issue has been highlighted by Goodhart (1975) who argues that “when a measure becomes a target, it ceases to be a good measure”. Evidence shows that incentivised performance indicators often lead to attempts to the game the system. In some cases, it can therefore be desirable to develop entirely forward looking performance monitoring systems that use indicators only to improve future policies and programmes rather than reward past performance.

Background information is necessary to understand an indicator

Each indicator should be accompanied by background information that allows an appropriate interpretation of it. The background information should specify the unit of measurement and describe the sample to which the indicator applies. If an indicator is subject to particular caveats or limitations, for example due to the nature of data collection, it should be mentioned, too. In addition to technical aspects, the policy objective for the indicator should be stated along with a brief explanation of why the objective is important. In some cases, an explanation why an indicator is a good measure for an objective can be helpful for readers who are not familiar with a subject.

The background information can also be used to describe the position of the indicator in the hierarchy of objectives. Furthermore, potential complementarities with other indicators could be mentioned. If several indicators are in a complex relation with each other, it can be helpful to represent them schematically in a graph such as the one in Figure 3.

Box 1 provides an example of background information that could be provided for each indicator.

Box 1. Example of background information provided with an indicator

This box provides a hypothetical example of the background information that could be provided with every indicator. Depending on the intended audience, some aspects could be elaborated in more detail, whereas others could be omitted. For example, if the background information was primarily intended for the general public, more emphasis could be put on why the outcome is important. If it was primarily intended for experts, more information regarding the data collection methods could be provided.

Indicator: Indicator **A1**: Share of children aged 6 who have at least satisfying language skills.

Corresponding policy objective: Objective **A**: Ensure a high-quality pre-school education in order to improve the schooling success of disadvantaged children.

Current value / target value: 90.3% / 95.0% by 2020.

Development over the past three years: ↗ (**+1.4** percentage points)

Unit of measurement: Percent of children at the age of six that attend a primary school and perform at least satisfactory in the standardised language comprehension test.

Outcome contributes to: Objective **B**: Reducing the share of early school leavers, Objective **C**: Improving the education of disadvantaged groups to foster social inclusion.

Outcome is influenced by: Objective **D**: Increasing the share of vulnerable children enrolled in pre-school education.

Why is it important to measure the outcome: Sufficient language skills are an important determinant of overall school success. Students without adequate language skills experience learning difficulties across many subjects that are hard to remedy at a later point in time.

Caveats: The indicator is based on a representative sample of 5 000 children from 103 schools and has a standard error of 1.2.

Source: Authors' elaboration.

Indicators can be tools to inform the public and ensure transparency

Indicators can be an important tool to convey key facts to the public and to form the foundation of an informed public debate. When they are well-designed, they contain information about fundamental conditions within a policy area in an easily comprehensible number. This makes them ideally suited for communication purposes. They can be used to focus public attention on issues and to showcase government initiatives. Furthermore, publishing them regularly increases transparency and accountability of the administration by providing clear yardsticks of what has been achieved.

In order to use their potential, indicators should be published regularly and be made easily accessible to other user from within the government and the public. Indicators can be made more comprehensible by colour-coding them according to how they have developed. A good example of how colour-coding can improve accessibility is the United Kingdom's "Biodiversity Indicators in Your Pocket (BIYP)" publication and its corresponding website.[13] However, colour-coding also creates incentives to manipulate the indicators to achieve better "colours". Therefore, it should only be used after careful consideration of the associated advantages and disadvantages.

INDICATORS AS STEERING TOOLS ON THE SUB-NATIONAL LEVEL

So far, the discussion has implicitly assumed that indicators are used on a national level. While all the issues that were mentioned so far also apply to indicators on the sub-national level, further aspects need to be considered. In particular, the role that indicators can play in encouraging cooperation among different sub-national governments should be taken into account.

Sub-national indicators can connect regional policies to national goals

Sub-national indicators can help policy makers to better understand how sub-national policies work, just as national indicators can improve the understanding of national policies. Furthermore, sub-national indicators are valuable for national policy makers because they can show how sub-national policies contribute to national objectives. For example, a local initiative to improve public transport might contribute to the national objective of reducing greenhouse gas emissions. National-level indicators cannot monitor the effects of many such local policies because individually they are too small to be captured by national indicators. Nevertheless, local policies provide important contributions to achieving national goals.

The effectiveness of local policies in contributing to national objectives can be monitored by sub-national indicators. As policies are implemented at different sub-national scales, indicators at corresponding scales have to be developed to monitor their performance. Ideally, national and sub-national policies are monitored by coherent indicators. While it will never be possible to distinguish exactly between the effects of national and sub-national policies, a set of comparable indicators at different national and sub-national levels contributes to obtaining a better understanding of the likely contributions.

Using sub-national indicators to monitor regionally differing effects of national policies

National policies do not affect all parts of a country equally. Depending on existing conditions, the same policies can cause different outcomes in different regions. A change in the taxation of rental property, for example, is likely to have much stronger effects on urban areas were renting is more common than in the rural countryside were most people own their own houses. Only when such regional differences in outcomes are taken into account is it possible to gain a thorough understanding of the consequences of a policy.

A single indicator for the entire country provides only the national average of an outcome and neglects the regional differences in outcomes. Often, this means it does not capture valuable information. [14]

[13] Department for Environment, Food and Rural Affairs (2014), available at: http://jncc.defra.gov.uk/page-4229.

[14] For example, regional differences in well-being are typically large within a single country. See OECD (2014) in this context.

In cases in which policies have very different effects in different regions, national indicators might even be not informative at all. If a policy causes an outcome to improve in half the country and worsen by the same magnitude in the other half of the country, a national indicator would show zero change. Clearly, this would not be an adequate description of the outcome of the policy.

For such policies, it is important to use sub-national indicators to monitor them appropriately. The choice of the sub-national level should be determined by the degree of regional variation in the effects of a policy. Ideally, an indicator should be defined for a region for which a policy can be expected to have relatively homogenous effects. In practice, this is not always possible because some policies have very local effects that vary from neighbourhood to neighbourhood.

Using indicators to achieve better policy coordination

Frequently, coordination gaps occur on sub-national levels of government. Sub-national governments are not cooperating as much with each other as would be socially optimal. Thus, their policies are less effective than they could be. Outcome indicators can contribute to greater coordination on the subnational level by emphasising the need for effective policies. By highlighting shortcomings that are the

Indicators vary in terms of how strongly they encourage coordination because not every outcome is equally affected by coordination between sub-national governments. This can be illustrated with an example from the public transport sector. Whenever public transport is provided locally, its quality is strongly affected by the degree of coordination between neighbouring municipalities. However, not every reasonable outcome indicator takes this into account. The share of residents that lives within 10 minutes walking distance to a bus stop is a meaningful outcome indicator that reflects an important quality of the public transport system. However, it is easily possible for a municipality to develop a bus network that performs well along this dimension, while at the same time offering poor connectivity to neighbouring municipalities. Thus, the indicator reflects only partially the objective of a typical public transport policy. In contrast, the number of people using the bus network is also a good indicator for the quality of the public transport system, but it has the additional advantage of being strongly affected by the degree of coordination between municipalities. If two municipalities coordinate their bus networks, it is unlikely to affect the first indicator but will probably have strong effects on the second indicator. Thus, the second indicator reflects the gains from coordination between municipalities better than the first one.

Often, indicators that measure objectives at intermediate levels are more likely to reflect the gains from coordination than those that measure objectives at lower levels. In the example above, the share of people using the bus network is partially determined by the share of people living within a certain distance of a bus stop; thus, the former is a higher level objective than the latter. Objectives at very low levels sometimes refer to very local outcomes that are not always affected by coordination. Therefore, it is important to measure outcomes at a sufficiently high level in order to encourage coordination.

Scoreboard indicators on the sub-national level

Scoreboard indicators can be used to monitor outcomes on sub-nationals levels, but some additional caveats have to be taken into account. The smaller the sub-national unit to which an indicator refers, the stronger it is affected by random fluctuations. This is especially important for indicators that refer to rare outcomes. An example for this could be the murder rate as a potential scoreboard indicator for crime. Most OECD countries have a murder rate in the very low single-digits per 100 000 inhabitants. Measured at the country level, the murder rate provides a fairly reliable average that fluctuates little from year to year. If it were measured at small sub-national units that have on average 10 000 inhabitants, the murder rate would be a highly unreliable measure that is very strongly influenced by random noise. Most sub-national units would report a murder rate of 0, whereas the few where one murder occurred would report a rate of

roughly 10 per 100 000 inhabitants.[15] Obviously, such fluctuations are primarily the result of random events that are hardly related to the actual situation in a region. Thus, indicators measuring such rare outcomes are not suitable for use at strongly disaggregated regional levels.

In some instances, an indicator might measure an outcome that is inherently heterogeneous across a country, for example because a public service can only be provided at certain locations. In these cases, the indicator should only be used for sufficiently high sub-national levels. Every sub-national unit should be large enough to be representative of the entire country with respect to the outcome in question. This can be illustrated by an example. The share of people aged 25 with at least a bachelor's degree could potentially be a scoreboard indicator for the state of higher education. It is likely that this share is much higher in cities that have a university than in cities that do not have a university. Disaggregating the indicator to sub-national units that are so small that some of them do not contain a university could therefore create misleading results. These sub-national units would likely have much fewer graduates than those that contain a university. The strong difference in the outcome could be perceived as a sign of undesirable regional disparities, even though it is generally accepted that higher education can only be provided at selected locations.

Lastly, it is important to interpret and attribute regionally disaggregated indicators appropriately according to regional responsibilities for the outcome. A region might have very high levels of carbon dioxide emissions because it is a centre of heavy industries that produce building materials such as steel and concrete. If these materials are used throughout the entire country, it would be misleading to attribute the carbon emissions that occur during their production only to the region where they are produced. The materials are essential for the functioning of the economy of the country and the carbon emissions that occur during production should be attributed to regions according to their use of the materials. Essentially, the responsibility for these emissions should be attributed to the consumer-region instead of the producer-region.

Problems of attributing outcomes correctly to regions could also occur in other situations. For example, patents might be developed at research centres but registered under the address of company headquarters that are located in different regions. In this case, it would be important to attribute the patents to the location of the research centre and not to the location of the company headquarter, as it is often done in practice.

INDICATORS IN INTER-GOVERNMENTAL CONTRACTS

This section discusses the use of outcome indicators in contracts between governments (in the following called outcome-based contracts). It focuses on aspects related to the use of indicators and touches only peripherally on many of the broader aspects related to the use of contracts between governments.[16]

Outcome-based contracts can be important tools for the public sector to manage interactions between different governmental actors. If they are well-designed and used at the right scale, they have several advantages. Most importantly, they provide incentives to the contractor to focus on the outcome instead of

[15] The statistical murder rate per 100 000 inhabitants in a region of 10 000 inhabitants is 10 if a single murder occurs in the region (1 murder / 10 000 inhabitants = 10 murders / 100 000 inhabitants).

[16] See for example OECD (2007) on inter-governmental contracts and Steffensen (2011) on performance based grants for much broader discussions of the use of contracts by governments.

just producing an output. Furthermore, they are useful for delegating responsibility regarding particular outcomes without the need of specifying in detail how to achieve them, thus giving the contractor greater responsibilities.

Despite these advantages, the use of outcome-based contracts in the public sector is still not widespread.[17] One possible reason for the slow uptake of outcome-based contracts is the challenges involved in designing them. In particular, it is difficult to set-up contracts that define outcomes appropriately and do not provide misaligned incentives. As outcome-based contracts give more freedom to the contractor, they also offer greater scope for undesirable actions and might inadvertently encourage the contractor to take such actions. Thus, outcome-based contracts need to be specified in a way that discourages the contractor from taking such actions. This paper highlights the potential advantages of outcome-based contracts, but nevertheless advocates for a cautious use. In general, it is more complex to develop appropriate outcome-based contracts than to use output-based contracts. Not in every case is the extra effort justified by the potential benefits from using them. This problem is compounded by the relative novelty of outcome-based contracts, which implies that governments frequently lack experience with their use.

Types of inter-governmental contracts

OECD (2007) defines two types of inter-governmental contracts; *transactional contracts* and *relational contracts*. Transactional contracts between governments resemble private sector contracts and generally aim to be complete. They try to describe all details of the transactional relation and state the required actions by the contracting parties. Implicitly, they assume that the involved parties primarily focus on their own gain and have little or no interest in cooperation beyond the scope of the contract. In contrast, the primary purpose of relational contracts is not to fix the parties into a complete set of enforceable rights and duties but to serve as mechanism for collective decision-making which generates trust and facilitates co-operation and information-sharing.

Relational contracts are often used in situations that are characterised by an absence of competition. Levels of government or governmental agencies often need to cooperate with each other and cannot simply find a different contractual partner for a task. In such situations, it is not possible to walk away from future contractual arrangements. Therefore, the contractual partners have to establish a long term relationship with each other that prevents conflicts from occurring. Although relational contracts may appear weak given the lack of emphasis on binding and enforceable constraints, OECD (2007) finds that they can help sustain a dynamic of co-operation among levels of government over the long run.

Outcome indicators in relational contracts

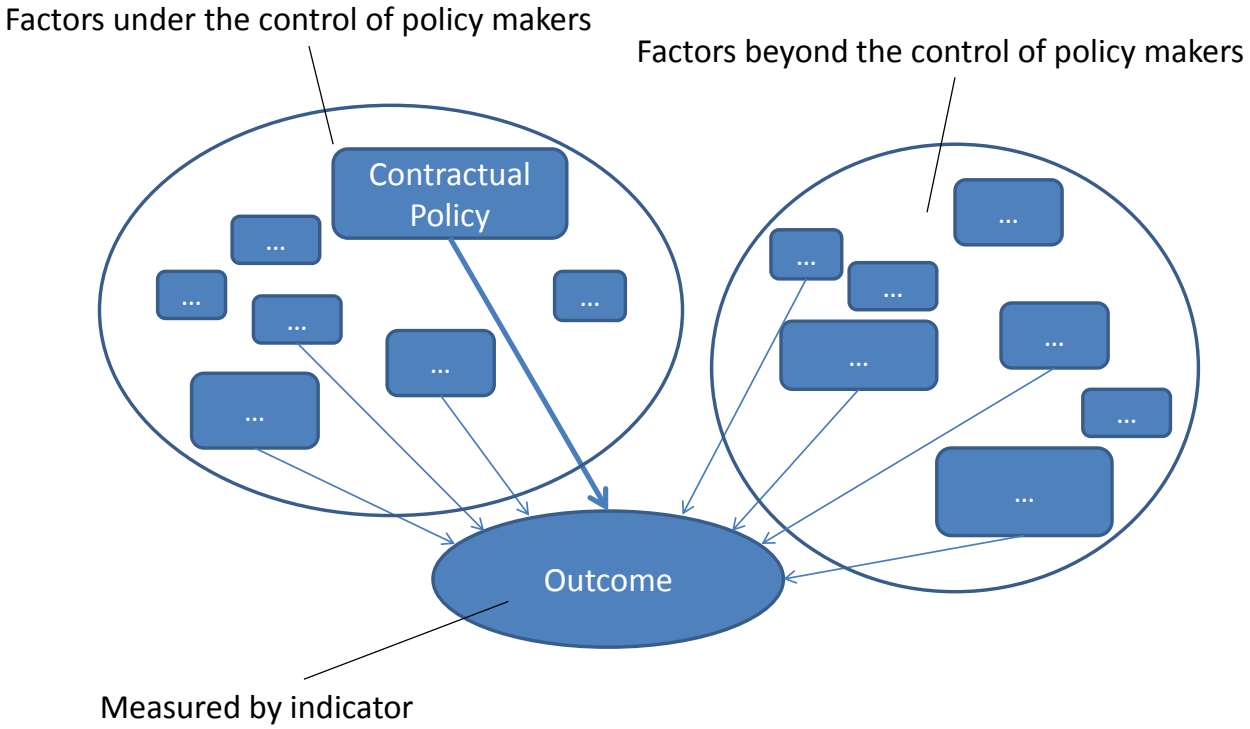
Outcome indicators are generally better suited for use in relational contracts than in transactional contracts. Relational contracts make it possible to take factors beyond the control of policy makers into account and adjust the contract if unforeseen circumstances occur. This is important when outcome indicators are used in contracts, because outcomes are only under partial control of policy makers.

One of the major problems of outcome-based contracts relates to the uncertainty of outcomes. As discussed earlier it is never possible to predict perfectly what the consequences of a policy will be. Sometimes, seemingly appropriate policies do not work in practice and in other cases minor policy reforms have unexpectedly large effects. Whereas it is usually possible to connect an input and an output directly to an individual policy, governments have only indirect control over outcomes. Policy makers can try to

[17] See Ng et al. (2009) for a general discussion of outcome-based contracts and an examination of their use by the British Ministry of Defence (MoD).

achieve a particular outcome but they cannot promise with certainty that they will achieve it even if they implement appropriate policies. This is the case mainly because outcomes are always influenced by many factors. Some of them are within the control of policy makers and some are beyond their control. Figure 4 illustrates this with a schematic graph.

Figure 4. Determinants of policy outcomes



Source: Authors' elaboration.

In light of this uncertainty, it is obvious that it is easier to use outcomes indicators in relational contracts than in transactional contracts, which offer less flexibility. Relational contracts are based on mutual trust instead of strict adherence to pre-agreed rules. Thus, contracting parties in relational contracts will find it easier to adjust the contract if outcome indicators changed due to factors beyond the control of policy makers.

Typically, the higher the level of the outcome, the more confounding factors influence it. This implies that many high-level outcome indicators are not well suited to be used as yardsticks even in relational contracts, because they lack the responsiveness to an individual policy. The resulting lack of precision of the indicator regarding the effectiveness of the policy makes it problematic to base a contract on it.

In contrast, it might be preferable to use outcomes that are closely connected to individual policies, as policy makers have a greater degree of control over them. For example, if a contract is supposed delegate responsibility for the efficient use of water to another government agency, an outcome such as *share of water lost due to leakage in the network* would be better suitable than an outcome such as the *water level in the main reservoir*. Both are valid outcome measures that can be influenced by public policy and could serve as important outcome indicators. However, the first measure is affected by fewer external factors than the second measure, which is heavily influenced by rainfall levels. Therefore, the first measure can be much more closely controlled by policy makers than the second one.

Importantly, implementing outcome-based contracts in the public sector is not only a question of setting up an appropriate contract. It often includes a transfer of the responsibility for an outcome from one governmental actor to another. Thus, not only responsibilities have to be transferred, but also that the government agency tasked with achieving the outcome has the necessary means to achieve them. This includes the legal rights to take the necessary measures, but also the availability of appropriate financial resources and the existence of sufficient institutional capacity to deal with the new tasks. When done successfully, this can contribute to capacity building at the local level and allow for more effective cooperation between levels of government.

Outcome indicators in transactional contracts

Transactional contracts between governmental actors are based on the assumption that any cooperation between the actors extends only to the terms specified in the contract. Compared to relational contracts, transactional contracts are more similar to typical private sector contracts. Because transactional contracts assume essentially non-cooperative behaviour, the role of incentives needs to be more closely scrutinised when outcomes are used in them.

An example can show how easily an outcome-based transactional contract could provide misaligned incentives. A contract between a regional government and a semi-independent governmental agency to provide training to unemployed individuals could be based on an outcome, such as the share of individuals that find a job within six months after completing the training. The payment to the governmental agency could be made dependent on the outcome; the higher the share of employed individuals is after six months, the higher is its payment.

This is a typical outcome-based transactional contract. However, without further provisions, the contract would most likely have undesired consequences. As it is based on the share of people who find a job, it provides incentives for the agency to select only those people for trainings that are likely to receive a job offer for training. As a first step, such a contract would therefore have to specify who will be eligible for the training. However, even then, there might be other possibilities for the agency to influence the selection of candidates. For example, it could schedule the courses for people it deems less likely to find a job at inconvenient locations or dates. This might induce them not to sign-up for the training. Alternatively, the agency could push participants to apply for jobs that are easier to get but are badly matched to their skills. For example, it could encourage highly educated individuals to apply for low-skilled jobs if these positions are easier to find. Similarly, the agency could push participants to apply for temporary jobs that do not offer a long-term perspective but will have participants employed at the evaluation date. While all actions would be beneficial for the agency because they increase its remuneration, none is desirable from the perspective of the government.

If transactional contracts between levels of governments are used and it is not possible to set up an outcome-based contract that rules out these and similar actions, it is preferable to use a more traditional output-based contract. In the case of the example, such a contract would simply specify that the agency has to provide training to a given number of participants and specify certain details. It would not refer to any outcome, such as the number of participants who find jobs. The contract would forego the use of a desirable incentive (i.e. to provide a training that is as effective as possible in helping people to find jobs) but would also not be affected by the undesirable incentives mentioned above.

Compared to outcome-based transactional contracts, contracts based on outputs are easier to develop and generally have lower risks of setting misaligned incentives. Nevertheless, good reasons exist to use outcome-based contracts. Whereas output-based contracts only provide incentives to deliver the contractually agreed output, outcome-based contracts create incentives to achieve the outcome. In other words, they provide incentives to achieve the goals that are the rationale for the contract. In many cases,

the contractor is better informed how to achieve an outcome than the client (i.e. the government issuing the contract). An outcome-based contract gives the contractor the opportunity to use the methods that work best to achieve the objective (in contrast to simply producing an output no matter its effectiveness in achieving the outcome).

In order to increase the use of outcome based contracts, relational elements in contracts can be strengthened. Many contracts between governments are neither purely transactional nor purely relational but include elements of both types. Fostering forms of cooperation between governments that strengthen relational elements in contracts can enable also the use of outcome indicators.

INDICATORS IN THE CONTEXT OF EU COHESION POLICY

The European Commission actively supports systematic monitoring and evaluation of policies. For the programming period 2014 – 2020, it has introduced a results-based imperative in its cohesion policy framework and requires that the outcomes of supported policies are monitored. Compared to earlier monitoring requirements, the new focus on outcomes represents a shift in the stance of the European Commission. For many EU member states, this made it necessary to create new monitoring frameworks or significantly modify their existing monitoring frameworks. This section discusses the particularities of policy monitoring under EU cohesion policy, in particular with respect to the varying monitoring requirements of the different programmes.

Aligning different monitoring frameworks

The monitoring requirements of the European Commission vary depending on the programme under which policies are support monitoring requirements vary. This is not an uncommon situation. Often, monitoring frameworks are introduced for varying reason and by different actors. Sometimes, they are introduced by national governments and cover most policy fields. In other cases, they refer only to specific policy areas or cover only some sub-national regions. Often, the different monitoring frameworks serve different purposes and use different methodologies.

Although such varying approaches are often justified, it is nevertheless desirable to ensure a minimum level of comparability across different performance monitoring frameworks. Besides allowing the comparison of different policy fields with each other, ensuring a minimum level of comparability also makes it easier to communicate the different monitoring frameworks.

The European Commission specifies that outcome indicators used to monitor programmes under the European Structural Fund (ESF) can be monitored by programme specific outcome indicators.[18] These are indicators that concern very low-level objectives. Typically, they refer to the share of participants in a programme for whom a desired outcome has occurred. Such an indicator could for example be the share of unemployed participants in a training programme that found a new job within six months of participating in the programme.

[18] See the guidance document on monitoring and evaluation of the European Cohesion Policy for the European Social Fund (European Commission (2014a), available at: ec.europa.eu/social/BlobServlet?docId=7884&langId=en).

In contrast, indicators for the European Regional Development Fund (ERDF) and Cohesion Fund (CF) refer to outcomes on a macro level.[19] They are supposed to measure the impact of a programme not just on those individuals or businesses affected by it, but on the entire population the programme is targeted at. For example, a typical ERDF indicator would not look at the share of youths who found a job after attending a training programme targeted at unemployed youths, but at the youth unemployment rate in general.

In addition to the different interpretations that the different types of outcome indicators have, they also have different data requirements. Typical outcome indicators for the ESF only require information about the group of individuals (or businesses, organisations, etc.) that is affected by a programme. This information is usually easy to collect and quickly available. In contrast, typical indicators for the ERDF require information for all individuals (or businesses, organisations, etc.) within the country, not just those that are affected by a programme. Such data is more difficult to collect and it typically involves longer time lags until it becomes available.

In order to prevent misinterpretations, the structural differences between the different types of indicators should be highlighted. If possible, it could also be desirable to collect two sets of indicators for key objectives; one on the programme level and one on the macro level. However, before doing so, it should be ensured that the gain from two different sets of indicators justifies the associated costs.

Not only do the requirements for indicators under the ESF and ERDF differ from each other, they are often very different from national performance monitoring and evaluation frameworks. In many cases, the monitoring and evaluation requirements for EU funded programmes are stricter than those practiced nationally. The monitoring requirements of the European Commission regarding outcome indicators emphasise the importance of ensuring that policies achieve the desired outcomes. This focus is important not just for programmes funded by the EU, but should be the underlying principle of every policy, no matter whether it is funded nationally or by EU sources. The monitoring requirements of the European Commission could be used as an occasion to modernise and extend national performance monitoring frameworks. In this context, experiences from the introduction of outcome indicators for EU programmes could provide valuable lessons how to design a broad monitoring framework based on outcomes.

[19] See the guidance document on monitoring and evaluation for the European Regional Development Fund and the Cohesion Fund (European Commission (2014b), available at: http://ec.europa.eu/regional_policy/sources/docoffic/2014/working/wd_2014_en.pdf).

CONCLUSION

This paper has discussed the design and use of indicators for policy monitoring. First, the paper argues that it is never possible to know with certainty in advance what the effects of a policy will be. From this, the need for policy monitoring follows directly. Without monitoring, effective policies cannot be distinguished from ineffective ones. Monitoring helps policy makers to identify ineffective policies and to change them. Second, the paper draws attention to the need for clear policy objectives. Only if objectives exist is it possible to judge whether policies achieve what they are supposed to achieve.

Three different types of indicators can be distinguished; input, output and outcome indicators. All should play a role in a monitoring framework, but it is important to use them according to their purpose. In particular, it is crucial to distinguish between outputs and outcomes and not to use output indicators to measure outcomes. From a policy maker's perspective, outcome indicators are arguably the most important indicators, as outcomes are the reason that a policy is implemented in the first place. If they are well-designed, outcome indicators have several advantages over other monitoring techniques. They provide regular, timely, and unambiguous feedback to policy makers. They can help policy makers to change course if policies are not working, can foster learning and capacity building and create transparency and accountability.

This paper has focussed on the design and use of indicators. However, it is equally important that information provided by indicators is translated into policies. This requires a framework to regularly analyse the indicators, to communicate the insights to all relevant authorities and most importantly to act upon them. Furthermore, it should include strategies on how the information from the indicators can be used for learning and capacity building and include provisions how to communicate the information gained from the indicators.

The purpose of indicators can be summed up using the terminology of this paper. They are outputs that are created to achieve the outcome of better policies. As is the case for all outputs, it is not guaranteed that indicators will achieve the outcome they are supposed to achieve. Only if they are of good quality and embedded in a framework that translates their insights into reforms can they contribute to better policies.

BIBLIOGRAPHY

Bartolini, D. (2013), "The Role of Incentives in Co-operation Failures", *OECD Regional Development Working Papers*, No. 2013/09, OECD Publishing, http://dx.doi.org/10.1787/5k46bj4t4nxs-en.

Department for Environment, Food and Rural Affairs (2014), UK Biodiversity Indicators 2014, London, available at: http://jncc.defra.gov.uk/pdf/UKBI2014.pdf.

Employment and Social Development Canada (2014), *Indicators of Well-being Canada* (data set), http://www4.hrsdc.gc.ca/.

EPA (2011), *Guide to Sustainable Transportation Performance Measures*, Environmental Protection Agency, available at http://www.epa.gov/smartgrowth/transpo_performance.htm.

European Commission (2014a), "Programming Period 2014-2020, Monitoring and Evaluation of European Cohesion Policy, European Social Fund, Guidance document".

European Commission (2014b), "The Programming Period 2014-2020, Guidance Document on Monitoring and Evaluation − European Regional Development Fund and Cohesion Fund – Concepts and Recommendations", available at: http://ec.europa.eu/regional_policy/sources/docoffic/2014/working/wd_2014_en.pdf.

European Commission (2014c), *Innovation Union Scoreboard 2014*, available at: http://ec.europa.eu/enterprise/policies/innovation/files/ius/ius-2014_en.pdf.

Goodhart, C. (1975), "Monetary Relationships: A View from Threadneedle Street", *Papers in Monetary Economics*, Vol. 1, Reserve Bank of Australia, Sydney.

OECD (2014), H*ow's Life in Your Region?: Measuring Regional and Local Well-being for Policy Making*, OECD Publishing, http://dx.doi.org/10.1787/9789264217416-en.

OECD (2013a), *OECD Regions at a Glance 2013*, OECD Publishing, Paris, http://dx.doi.org/10.1787/reg_glance-2013-en.

OECD (2013b), *Government at a Glance 2013*, OECD Publishing, http://dx.doi.org/10.1787/gov_glance-2013-en.

OECD (2012a), *OECD Environmental Outlook to 2050: The Consequences of Inaction*, OECD Publishing, http://dx.doi.org/10.1787/9789264122246-en.

OECD (2012b), *Settling In: OECD Indicators of Immigrant Integration 2012*, OECD Publishing, http://dx.doi.org/10.1787/9789264171534-en.

OECD (2011), *OECD Guide to Measuring the Information Society 2011*, OECD Publishing, http://dx.doi.org/10.1787/9789264113541-en.

OECD (2009), *Governing Regional Development Policy: the use of Performance Indicators*, OECD Publishing, http://dx.doi.org/10.1787/9789264056299-en.

OECD (2007), *Linking Regions and Central Governments: Contracts for Regional Development*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264008755-en.

OECD/European Union/JRC (2008), *Handbook on Constructing Composite Indicators: Methodology and User Guide*, OECD Publishing, http://dx.doi.org/10.1787/9789264043466-en.

Kellogg Foundation, W. K. (2006), *Using logic models to bring together planning, evaluation, and action: Logic model development guide*, https://www.wkkf.org/resource-directory/resource/2006/02/wk-kellogg-foundation-logic-model-development-guide.

McCann, P. (2015), "*The Regional and Urban Policy of the European Union: Cohesion, Results-Orientation and Smart Specialisation*", Edward Elgar, Cheltenham.

Oman, C. P. and C. Arndt (2010), "Measuring Governance", *OECD Development Centre Policy Briefs*, No. 39, OECD Publishing, http://dx.doi.org/10.1787/5km5z9nnxwxw-en.

Steffensen, J. (2010), "Performance-Based Grant Systems: Concept and International Experience", United Nations Capital Development Fund, http://www.uncdf.org/en/node/2304 (accessed November 10, 2015).

Umweltbundesamt (2010), *Establishment of an Indicator Concept for the German Strategy on Adaptation to Climate Change*, Federal Environment Agency, http://www.uba.de/uba-info-medien-e/4031.html.

Van Dooren, W., G. Bouckaert and J. Halligan (2015), *Performance management in the public sector* (2nd edition) Routledge: London.

World Bank (2013), *Independent Panel Review of the Doing Business report*, Washington, available at: http://www.dbrpanel.org/sites/dbrpanel/files/doing-business-review-panel-report.pdf.