



# Toolkit for Evidence-Informed Policymaking in Migrant Integration

## Section 3 Evaluation to Strengthen Ongoing and Future Integration Programmes

By Jasmijn Slootjes and Maria Belen Zanzuchi

January 2023

The full toolkit can be found at:  
[www.migrationpolicy.org/research/toolkit-evidence-policymaking](http://www.migrationpolicy.org/research/toolkit-evidence-policymaking)



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004635.



Title	Toolkit for Evidence-Informed Policymaking in Migrant Integration
Responsible Authors	Jasmijn Sloopjes and Maria Belen Zanzuchi, Migration Policy Institute Europe
Version	Final

**SPRING** is a EU-funded project focusing on the integration of recently arrived migrants in the context of the large-scale arrivals of refugees and other migrants since 2014. It aims to develop a toolbox to improve the innovation, effectiveness and sustainability of the work done by Europe's integration stakeholders at national, regional and local levels. The project mobilises significant research, networks and communications capacity and gathers, summarises and shares the best available research and evidence on the effectiveness, innovation, transferability, sustainability and evaluation methods for integration policies and practice.

**The SPRING Platform [integrationpractices.eu](https://integrationpractices.eu) is the main hub to make the project results available to practitioners as well as to the general public.**

## Acknowledgements

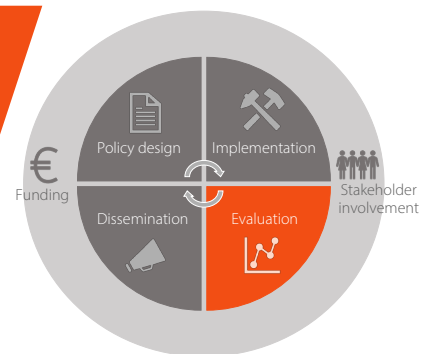


The SPRING project received funding from the European Union's Horizon 2020 research and innovation programme under the grant agreement 101004635. The contents of this document are the sole responsibility of the authors and do not necessarily reflect the views of the European Union. The European Commission and the European Research Executive Agency are not responsible for any use that may be made of the information it contains.

The authors thank their Migration Policy Institute (MPI) colleagues Meghan Benton for her insightful comments and feedback that played a crucial role in shaping this toolkit and Lauren Shaw for the excellent edits. They also thank MPI Europe interns Ilaria Marconi, Johanna von Kietzell, Giovanni Penna, and Vilde Hefte for their invaluable research assistance. In addition, the authors are grateful for the support they received from their colleagues in the SPRING Consortium. The authors are particularly grateful for Guia Gilardoni's excellent leadership throughout the SPRING project, the interesting discussions with Peter Scholten about improving the research-policy-practice nexus, and the valuable insights from Alexander Wolffhardt and Sinem Yilmaz about the transferability of evidence.

Critical input on the design of this toolkit also came from the experts who participated in the March 2022 roundtable 'Implementing What Works: Promoting Evidence-Based Migrant Integration Policymaking' and the September 2022 workshop on the role of communities of practice and cooperation networks in supporting more inclusive and evidence-informed integration policy and practice, both organised as part of the SPRING project. Finally, the authors thank the following interviewees for their valuable insights: Helen Baillot, Peter De Cuyper, Renée Frissen, Jenny Phillimore, Meta Van der Linden, and Jürgen Wander.

MPI Europe is an independent, nonpartisan policy research organisation that adheres to the highest standard of rigour and integrity in its work. All analysis, recommendations, and policy ideas advanced by MPI Europe are solely determined by its researchers.



### 3. Evaluation to Strengthen Ongoing and Future Integration Programmes

#### Key takeaways

- ★ Evaluation—the structured assessment of a policy’s process, impact, and/or cost-effectiveness—improves policy outcomes, promotes good governance, helps save on costs, increases accountability, and aids with policy learning.
- ★ Decisions about which type of evaluation to use should be based on the evaluation’s goal, the resources available, and the characteristics of the policy. Some of the most common types are:
  - *Process evaluations* focus on how a policy was implemented. They identify any problems that were encountered and how they were solved, the resources used to deliver services (inputs), the type and quantity of services delivered (outputs), and the beneficiaries of those services.
  - *Impact evaluations* assess a policy’s effects, for example on integration outcomes. Doing so is complicated because it can be challenging or at times impossible to say what would have happened without the policy and to establish a direct causal link between a policy and an observed integration outcome, given other factors at play. Randomised controlled trials, quasi-experimental techniques, and theory-based impact evaluations are examples of this type of evaluation.
  - *Value-for-money evaluations* systematically assess whether the money, time, or other resources (inputs) invested in an intervention represent good value. This includes cost-effectiveness and cost-benefit analyses.
- ★ Policymakers can leverage a range of existing tools and strategies to increase survey response rates among migrants, to overcome different forms of bias in how research is conducted and published, and to effectively commission an evaluation.

Evaluation is a structured and objective assessment of a policy, programme, or intervention. It is a crucial tool to understand whether policies are effective, for whom, and why. Evaluations create new evidence and help identify how policies can be improved and made more cost-effective.

### *In this section, you will learn...*

- why policy evaluation is important;
- what the different types of evaluations are, what the advantages and disadvantages of each are, and how to choose the best type for your work;
- how to overcome some of the methodological obstacles that can make doing research among migrant and refugee communities challenging; and
- how to effectively commission an evaluation.

## **3.1 What is evaluation and why should I evaluate?**

Evaluations differ in their scale and objectives, but all aim to answer questions<sup>10</sup> such as:

- Is the intervention or policy working as intended?
- What is its overall impact?
- Is it working differently for different groups, and why?
- If we were to do it again, how could the policy be improved?
- Is it good value for money, or could we achieve the same goals in a more cost-effective way?

Evaluation is *not*<sup>11</sup>...

- Spending reviews – studies of whether public interventions align with government priorities, with the goal of increasing available funds
- Monitoring – checking progress against established targets, with evidence reported in order to show that goals are achieved and resources are well spent
- Performance management – when an organisation involves its employees as contributors to the process of improving the organisation’s effectiveness
- Audit – a review of whether established criteria, such as compliance with financial and legal requirements, are reflected in information collected about an organisation’s operations

<sup>10</sup> This list is adapted from Section 1.2 of UK Treasury, *Magenta Book: Central Government Guidance on Evaluation* (London: UK Treasury, 2020).

<sup>11</sup> Organisation for Economic Cooperation and Development (OECD), *Improving Governance with Policy Evaluation: Lessons from Country Experiences* (Paris: OECD Publishing, 2020).

Done well, policy evaluation can improve policy outcomes, cost-effectiveness, policy learning, good governance, and decisionmaking. This holds benefits for government budgets, the lives of migrants and refugees, and the broader society. Some of the most important reasons to integrate evaluation into your work are summed up in Table 3.1.

TABLE 3.1

**Policy evaluation’s top benefits**

<i>Why evaluate?</i>
<b>Identify risks:</b> Policy evaluation helps identify dependencies, uncertainties, and risks. It also supports policy design and implementation adaptations to account for uncertainties and minimise risks.
<b>Facilitate accountability:</b> Evaluations hold policymakers, practitioners, and other stakeholders accountable to the objectives they set for themselves.
<b>Promote learning:</b> Evaluations generate lessons about what works and under which conditions, which stakeholders can learn from and integrate into future work.
<b>Respond to external scrutiny:</b> Evaluations provide support during official audits but also crucial information that can be shared with citizens, taxpayers, and electorates about how policies are performing.
<b>Improve stakeholder engagement and empowerment:</b> Involving stakeholders in the evaluation process, in addition to resulting in higher-quality evaluations, is an opportunity for policymakers to strengthen stakeholder engagement and empowerment. It builds a shared understanding of what the policy is and aims to do, and policymakers benefit from stakeholders’ expert knowledge.
<b>Boost cost-effectiveness:</b> Evaluations can identify whether policies provide value for money and point to ways to implement existing policies in a more efficient and cost-effective way, while achieving the same or better results.
<b>Increase effectiveness of service delivery and implementation:</b> Evaluations can improve the effectiveness of policy implementation and delivery of services by informing decisions about ongoing and future policies.
<b>Strengthen understanding of the context and conditions under which a policy works:</b> A policy may work for some groups and under specific conditions, but not in all cases. Evaluations help determine the context and conditions under which policies are most effective.
<b>Informing decisions about ongoing and future policies:</b> Evaluations play a key role in improving decisionmaking, for example about whether to discontinue, continue, or even scale up or expand a policy.

Sources: Michael Howlett, ‘Policy Analytical Capacity and Evidence-Based Policy-Making: Lessons from Canada’, *Canadian Public Administration* 52, no. 2 (2009): 153–175; Andrew Feldman and Rebecca Maynard, ‘The Evidence-Based Policy Revolution Waiting to Happen’, *Government Executive*, 30 January 2020; Pew Charitable Trusts and MacArthur Foundation, *Evidence-Based Policymaking: A Guide for Effective Government* (Washington, DC: Pew-MacArthur Results First Initiative, 2014).

### **3.2 How do I choose the evaluation approach right for my work?**

There are many different ways to approach evaluation. Each has advantages and disadvantages, and each will help you answer a different set of questions. Selecting the right evaluation approach—or combination of approaches—requires careful consideration of:

- the goals you hope to achieve (i.e., what questions you want to answer);

- the nature of the policy;
- the context in which the policy operates;
- the availability of data; and
- the resources and staff or external capacity available to support the evaluation.

Evaluation approaches, while there are many, can be grouped into the three broad categories shown in Figure 3.1, based on the goals they aim to achieve.<sup>12</sup> Figure 3.2 offers a set of questions and decision points that can help you think through which of these broad categories and which more specific evaluation types are right for a particular situation. Further information about different types of impact evaluations and value-for-money evaluations can be found in Sections 3.3 and 3.4.<sup>13</sup> Even if you never use these methods yourself, understanding the requirements and advantages of different approaches can help you to better understand policy evaluations conducted by others and to commission high-quality evaluations from external evaluators.

FIGURE 3.1

### Key characteristics of process, impact, and value-for-money evaluations



**Process evaluation** focuses on the implementation process to determine how it went and ways it could be improved.

*Key questions:*

- ✓ Is the policy being implemented as intended?
- ✓ Have any practical problems been encountered and how have they been addressed?
- ✓ What funding, staff, and other resources (inputs) are being used to implement the policy, and are they enough?
- ✓ What types and quantities of services have been delivered (outputs)?
- ✓ Who has benefitted from the services provided?
- ✓ What worked well and what did not? What could be improved?



**Impact evaluation** assesses the intended changes (objectives) and unintended changes caused by an intervention.

*Key questions:*

- ✓ Is the intervention effective in achieving the policy's goals?
- ✓ Did the intervention cause the change in outcomes, or would observed changes have occurred anyway without the intervention or as a result of other factors?
- ✓ Did the intervention result in unintended outcomes?



**Value-for-money (VfM) evaluation** systematically assesses whether an investment of money, time, or other resources (inputs) in an intervention represents good value by causing positive outcomes.

*Key questions:*

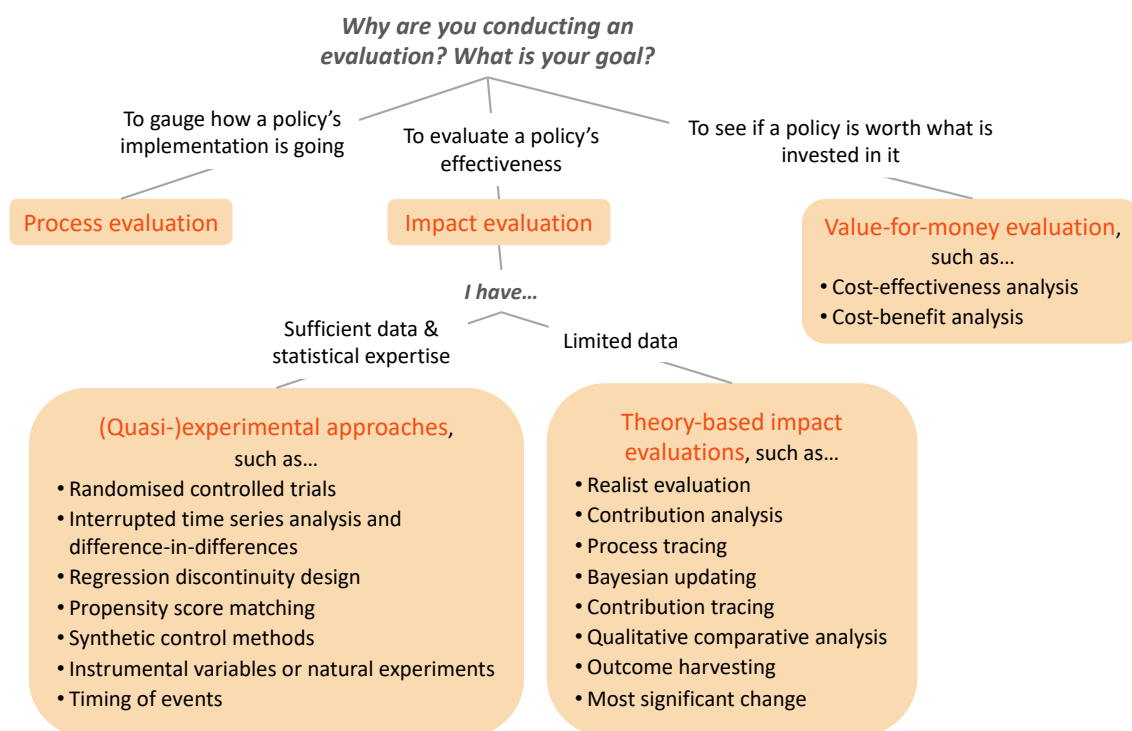
- ✓ Were resources used effectively?
- ✓ Is the intervention cost-effective, compared to alternative interventions or to doing nothing?
- ✓ Do the benefits outweigh the costs of the intervention?

<sup>12</sup> TSNE, 'Process Evaluation vs. Outcome Evaluation', updated 14 June 2018; UK Treasury, *Magenta Book: Central Government Guidance on Evaluation* (London: UK Treasury, 2020).

<sup>13</sup> There are no subtypes of process evaluation. For more information about how to set up a process evaluation, see the guidance provided by the Medical Research Council in Graham F. Moore et al., 'Process Evaluation of Complex Interventions: Medical Research Council Guidance', *BMC Public Health* 350 (2015): 1–7.

FIGURE 3.2

**Decision tree for choosing the right evaluation type**



**3.3 Assessing a policy’s effectiveness using impact evaluations**

Impact evaluation is complicated. It is methodologically challenging to determine whether a policy (for example, an integration course) has a causal effect on immigrants’ integration outcomes (for example, their labour market integration). There are different reasons why establishing impact is so difficult, including:

- **Unobserved counterfactuals:** When measuring the impact of a policy on an individual’s integration outcomes, it is impossible to know what the outcomes would have been without the policy.

*For example, in a study of the labour market integration outcomes of participants in an integration course, it is impossible to know what those individuals’ labour market outcomes would have been had they not attended the course.*

- **Spurious correlations:** Correlation is not the same as causation. More simply put, just because two things are related, that does not mean that one has caused the other.

*For example, if after six months 80 per cent of the participants in Integration Course A have found a job whereas 30 per cent of participants in Integration Course B have found a job, it would be incorrect to simply assume that Course A is more effective than Course B.*

There are two reasons for this:

- **Selection bias:** Ideally, participants for a study should be selected in a randomised way that creates either a sample representative of a broader population of interest (e.g., all refugee women) or a set of study groups that have roughly similar characteristics, making it possible to draw comparisons between their outcomes. Selection bias occurs when study participants are not randomly selected or allocated to different intervention and control groups, meaning they are not representative of the broader population or comparable to each other.

*For example, if the participants of the two courses had very different characteristics, it may not be possible to compare their outcomes. If those in Course A had a higher level of education and skills in in-demand fields, while those in Course B had high levels of illiteracy and work experience that did not match local labour market needs, Course B participants would be less likely to find a job than Course A participants no matter the quality of their integration courses.*

- **Confounding factors:** At first glance, two things (e.g., a policy and integration outcomes) may appear to be related, but what if there was a third thing affecting them both? That confounding factor artificially creates a correlation, even though there is no causal relation between the two.

*For example, Integration Course A was based in a large city with plenty of employment opportunities and job vacancies, while Integration Course B was based in a rural area with almost no employment opportunities and job vacancies. In this example, a migrant's location is the confounding factor because it determines both which integration course the migrant participates in and shapes the migrant's labour market integration outcomes due to differences in the availability of jobs.*

The process of determining whether there is a causal relation between, for example, a policy and integration outcomes is called **causal inference**. There is a wide—and growing—array of quantitative and qualitative research methods that can help you to establish evidence of a causal relationship. However, not all research methods create evidence of equal quality (see Section 2.4 for more on assessing the quality of evidence).

The rest of this section discusses research methods that can be used to establish causal impact: experimental methods, including randomised controlled trials; quasi-experimental methods; and theory-based impact evaluation.



**Want to learn more** about causal inference? Check out:

- ★ [This YouTube video](#), in which Professor Jens Hainmueller of Stanford University provides an overview of causal inference in migration studies. It was recorded at the Summer Institute in Migration Research Methods organised by the Berkeley Interdisciplinary Migration Initiative at UC Berkeley.



## **Experimental and quasi-experimental approaches**

**Experimental approaches** are the gold standard in evaluation. *Randomised controlled trials (RCTs)* are a type of experimental approach that randomly assigns individuals to either be exposed to a policy or not exposed to it, with the latter making up the so-called control group. Because individuals are assigned to groups randomly and not based on their characteristics, the groups should be similar. Any observed differences between the groups could therefore be interpreted as policy impact. In practice, it is often impossible to do RCTs or to randomly assign individuals to be affected—or not—by a certain policy.

**Quasi-experimental approaches** allow evaluators to mimic the conditions of an RCT and create an artificial control/comparison group so they can infer the impact of a policy on specific outcomes as accurately as possible. Common types of quasi-experimental studies include:

- **Interrupted time series analysis** tests whether launching, ending, or changing a policy leads to causal change in integration outcome trends over time. This method assumes that trends (for example, a steady increase in employment rates) would continue over time, unless affected by a change in policy. This method aims to identify policy impact by comparing a population’s integration outcomes following the introduction of a new policy to those that would have been expected if the existing trend were to continue as before. A *difference-in-differences* design further strengthens this type of study by comparing ongoing trends in a control group.<sup>14</sup>
- **Regression discontinuity design** compares the integration outcomes of individuals or groups just on either side of an arbitrary threshold that determines whether they are affected by a policy or programme, with those not affected acting as an artificial control group. For example, if a policy targets a specific age group (migrants older than age 35) or a specific geographic area (those in a specific municipality), a study using regression discontinuity design would compare the integration outcomes of migrants just below the threshold (those who are 34 years old or those living in a neighbouring municipality) with the outcomes of migrants who just made it into the group affected by the policy (those who are 35 years old or those who just barely live inside the municipality’s borders).<sup>15</sup>
- **Propensity score matching** creates a comparison group that is as similar as possible to the group being exposed to the policy (the intervention) by using a statistical technique that matches the intervention group with individuals not in the intervention group. For example, to study whether an integration course is effective, this type of study might compare the integration outcomes of migrants in the integration course to those of a comparison group composed of individuals who are as similar as possible to the course participants, as determined using available data on relevant characteristics. However, not all relevant characteristics may be used to create the comparison group, and there are often structural

14 For an example of a difference-in-differences design in migrant integration studies, see Jens Hainmueller and Dominik Hangartner, ‘Does Direct Democracy Hurt Immigrant Minorities? Evidence from Naturalization Decisions in Switzerland’, *American Journal of Political Science* 63, no. 3 (2019): 530–547.

15 An example of a Regression Discontinuity Design in migrant integration studies can be found here: Jens Hainmueller et al., ‘Protecting Unauthorized Immigrant Mothers Improves Their Children’s Mental Health’, *Science* 357, no. 6355 (2017): 1041–44.

differences between the group that is exposed to the policy and the group that is not not, which can make it difficult to use this method.<sup>16</sup>

Whether and which (quasi-)experimental approaches fit your work will depend on the availability of resources, the nature of the intervention, access to data, and other factors.

#### Advantages:

- Experimental approaches, and to a lesser extent quasi-experimental approaches, provide the most reliable way to assess whether policies are effective, and they create robust evidence.
- Impact is usually measured in a standardised and quantifiable way, which allows policymakers to more easily compare the effectiveness of different policies.

#### Disadvantages:

- Experimental and quasi-experimental approaches may not be possible, whether as a result of ethical concerns (e.g., about providing a potentially beneficial service to some people but not others in order to create a control group) or because it is simply not possible to create a control group.
- Experimental and quasi-experimental approaches are very resource intensive and require considerable statistical expertise, which may not be available in house.



**Want to learn more** about (quasi-)experimental approaches?

Check out:

- ★ The [Impact Evaluation in Practice](#) handbook by the World Bank (also available in Portuguese and Spanish) explains different experimental and quasi-experimental methods for assessing policy impact.
- ★ The [Magenta Book: Central Government Guidance on Evaluation](#), published by the UK government, provides an overview of experimental and quasi-experimental methods and their advantages and disadvantages (see page 48) and a decision tree to help you decide which method is the best option for your work.

<sup>16</sup> Jens Hainmueller and Dominik Hangartner, 'Who Gets a Swiss Passport? A Natural Experiment in Immigrant Discrimination', *American Political Science Review* 107, no. 1 (2013): 159-187.

## Theory-based impact evaluations

Theory-based approaches to impact evaluation focus on the questions of why and how policies are effective in causing changes to outcomes. Theory-based approaches often use a theory of change that explains how a policy is expected to produce certain outcomes. This theory of change includes a *logic model*—a hypothesised chain of causes and effects describing a sequence of inputs leading to outputs, and the mechanisms that translate inputs into outputs. Once developed, a theory of change is tested using evidence from multiple sources. This approach is often combined with other evaluation approaches because of its reliance on a wide range of evidence sources.

### Advantages:

- Theory-based approaches allow evaluators to explain why interventions are working or are not working.
- Theory-based approaches can be applied in all settings, unlike (quasi-)experimental methods.
- Theory-based approaches can be used as part of less resource-intensive evaluations and can add credibility to and increase the quality of these type of evaluations.

### Disadvantages:

- Theory-based approaches do not provide a quantitative measure of a policy’s impact on an observed outcome (effect size).
- In some cases theory-based approaches may end up being built on multiple theories of change, and evaluators may need to test and reconcile differences between them.



**Tip:** The [Ethnic and Migrant Minority Survey Question Data Bank](#) by Ethmig Survey Data is a collection of questionnaires you can use in quantitative surveys that target ethnic and migrant minority respondents.



**Want to learn more** about theory-based impact evaluations? Check out:

- ★ [Theory-Based Approaches to Evaluation: Concepts and Practices](#) by the Treasury Board of Canada Secretariat provides information on how to use theory-based approaches and their strengths and weaknesses.
- ★ [The Magenta Book: Central Government Guidance on Evaluation](#), published by the UK government, provides an overview of different types of theory-based impact evaluations (see page 45). These include realist evaluation, contribution analysis, process tracing, Bayesian updating, contribution tracing, qualitative comparative analysis, outcome harvesting, and most-significant change.

### **3.4 Assessing bang for your buck with value-for-money evaluations**

Value-for-money evaluation approaches can serve several purposes. They can help you make smart decisions that improve a policy’s efficiency, making it possible to reduce the amount of resources that need to be allocated while achieving the same or better impact. These approaches also make it possible to assess the broader financial impacts of a policy.

#### ***Cost-effectiveness analysis***

Cost-effectiveness approaches focus primarily on reducing the costs (input) of a policy, generally while achieving the same or a better effect (output). Often, this type of analysis compares the outcomes and the costs of two or more policies to gauge which provides a better return on investment. The policy outcomes have usually already been established by impact evaluations and should be measured broadly, including different indicators of integration (such as health, well-being, employment, sense of belonging, social contact with the receiving-society population). With many integration policymakers operating in a context of limited funding, cost-effectiveness analysis is an important tool in efforts to get the same bang for fewer bucks.

Advantages:

- Cost-effectiveness approaches help policymakers to use limited funds in a more efficient way.
- The outputs (for example, learning the host country’s language) do not have to be assigned a monetary value; instead, they can be measured in whatever way makes sense for the nature of the output (for example, increases in language proficiency levels), and then compared to the monetary costs of the policy. This makes this approach much easier to apply.

Disadvantages:

- Definitions of costs sometimes differ, and this may complicate comparisons between policies.
- What constitutes a one-unit increase in policy impact may not necessarily reflect a consistent amount of change. Taking the example of improvements in language proficiency, a language learner who advances from a very low to a moderate level will likely find it significantly easier to communicate in daily life, whereas a learner who advances from a moderate to a high level may not see as big of a difference in daily communication. Thus, the units of change (advancing one level) may not have the same amount of real-life impact.

## Cost-benefit analysis

Cost-benefit analysis is a tool to calculate the broader social value of investments in integration initiatives. This approach promotes a more holistic view of policy outcomes, including the broader financial and social impacts. A cost-benefit analysis approach would, for example, look at the broader impact of promoting language acquisition among migrants, including improved access to the labour market and therefore reduced dependence on social welfare, but also better social integration and stronger social cohesion in the broader society.

Cost-benefit analysis can follow two models:

**Predictive models** use economic modelling methods to predict the net social value of a policy or intervention by making assumptions about its impact based on existing evidence. Such models can predict how initial outcomes could have broader effects, even when there is not enough evidence on whether an intervention achieves its objectives.

**Evaluative models**, instead of predicting, assess actual costs and outcomes. Ideally, these models also look at broader effects such as indirect positive impact on the rest of the household when one household member finds a job. These broader effects are often hard to measure and assign a monetary value, and most evaluative models therefore focus solely on the costs and initial outcomes.

The two models can also be merged. Evaluations can be used to test prior predictions, for example by designing a feedback loop where a predictive model improves over time thanks to the input of better evidence. The models should include an estimate of uncertainty and be periodically reviewed.

Advantages:

- Cost-benefit analysis captures both short- and long-term impacts in a systematic way.
- Cost-benefit analysis takes a more holistic approach to defining the impacts of a policy, which can be helpful in a multifaceted policy area such as migrant integration.

Disadvantages:

- It can be challenging to assign a monetary value to the broader impacts of a policy.

### BOX 3.1

#### Case study: Cost-benefit analysis of refugee policy in Amsterdam

The City of Amsterdam, through an external consultancy firm (LPBL), carried out a cost-benefit analysis to evaluate and improve its 2016–18 Amsterdam Approach refugee integration programme. The cost-benefit analysis, carried out every six months, considered all extra costs for activities targeted to refugees, including client and programme management, language training, internships, and other activities. These costs were measured against all extra benefits, for instance a decrease in unemployment benefits, higher tax revenues, more long-term educational benefits, and better quality of life.

In the first year of programme implementation, costs and benefits were collected from a sample of 1,500 refugees participating in the programme. They were then compared with the results of a control group of 3,000 refugees who did not participate in the programme, collected from historical data. After one year in the programme, the employment of refugees was 15 per cent higher than for those in the control group. Programme participants also found jobs more quickly than the control group, with estimates that 50 per cent might not need unemployment benefits within three years of joining the programme. The programme's benefits were found to exceed its costs by 50 per cent in the most conservative estimate. For every euro invested, there would be a 1.50 euro gain.

Sources: Organisation for Economic Cooperation and Development (OECD), *Working Together for Local Integration of Migrants and Refugees in Amsterdam* (Paris: OECD Publishing, 2020); European Social Fund, 'The Amsterdam Approach Fast-Tracks New Citizens', updated 10 July 2019.



**Want to learn more** about cost-benefit analysis? Check out:

- ★ The Migration Policy Institute Europe report [A Needed Evidence Revolution: Using Cost-Benefit Analysis to Improve Refugee Integration Programming](#) provides an overview of how cost-benefit analysis can be used to improve integration policies and introduces innovative ways to measure integration outcomes.
- ★ The [Supporting Public Service Transformation: Cost-Benefit Analysis Guidance for Local Partnerships](#), a technical guide published by the UK Treasury, walks through each step of conducting a cost-benefit analysis, different types of costs, how to calculate net present budget impact, financial return on investment, and net present public value. It also provides examples of cost-benefit analyses conducted in the United Kingdom.

### **3.5 Overcoming methodological challenges when studying integration policy impacts**

Evaluating integration policies can be particularly challenging. Migrant and refugee communities may be hard to reach, linguistic barriers can hamper data collection, and integration is a long-term and complex process that can be difficult to measure. This section offers useful tools and best practices to help you overcome some of these methodological challenges.

#### ***Improving survey response rates***

High response rates are important to find out whether policies have an impact on integration outcomes and to prevent bias (see the section below). Yet, language barriers, time constraints, distrust of government, and other obstacles can result in poor response rates.<sup>17</sup>

Strategies for improving response rates include:<sup>18</sup>

- ✓ Explain to people who respond to the survey how it may help improve policies and how they will benefit (for example, by improving the delivery of a service they receive).
- ✓ Use easy-to-understand language instead of formal or academic language when addressing survey respondents, whether in writing or verbally over the phone/in person.
- ✓ Offer financial incentives to participate, such as a one-time small payment or randomly selecting some respondents to win a prize.
- ✓ Send reminders to potential participants encouraging them to complete the survey.
- ✓ Leverage the benefits of paper-based surveys, even as online surveys become more common. Not everyone has a computer, internet, and the necessary digital literacy to navigate an online survey, and handing out paper-based surveys in person can allow respondents to ask questions about its purpose and contents.
- ✓ Keep surveys short.
- ✓ Collaborate with community organisations and representatives to reach respondents who distrust government agencies and who would be less willing to fill in a survey coming directly from one.
- ✓ Meet hard-to-reach communities where they are, including at schools, local markets, community centres, religious institutions, and other usual places of gathering.

17 Joan Font and Mónica Méndez, eds., *Surveying Ethnic Minorities and Immigrant Populations: Methodological Challenges and Research Strategies* (Amsterdam: IMISCOE Research and Amsterdam University Press, 2013).

18 Gerhard Van de Bunt and Lorraine Nencel, *Social Research Methodology: Qualitative and Quantitative Approaches* (Amsterdam: VU University Amsterdam, 2011).

## Dealing with bias

**Selection bias** occurs in surveys when participants are not randomly selected from the population they represent. For example, using an online survey advertised via social media to collect data about integration among migrant groups will create a bias toward migrants who have access to the internet and a digital device and who use social media. Those without social media and/or without access to digital tools, groups that are disproportionately likely to be older or less well integrated, will not have a chance to participate. This skews the survey results because the population surveyed is not an accurate representation of the broader population of interest.

Strategies to prevent selection bias include:

- ✓ Randomly select study participants. Depending on the goals of the evaluation and practical constraints, this could involve simple random sampling (randomly selecting individuals in population records, often using computer-generated sample), stratified random sampling (taking a random sample from different predefined groups, for example groups of migrants), or cluster sampling (selecting clusters, for example a geographic area, with a probability method and selecting all individuals within a cluster to be part of the sample).<sup>19</sup>
- ✓ If random selection is not possible, reflect on how the selection method(s) used may bias the sample and try to mitigate this by diversifying selection and recruitment methods.
- ✓ After completing the evaluation, always critically reflect on how the selection of participants for the study may affect its results and include this information in the evaluation report's discussion of the results.

**Attrition bias** occurs in longitudinal studies (those conducted at multiple points in time) when some people who initially join a study later drop out. People usually do not drop out of a study at random; it is especially likely among those with limited interest and those facing poverty or other circumstances that can make participation difficult, but it can also occur if participants move, migrate, or change contact details, which may be more common among some demographic groups than others. The result of this non-random loss of some participants is a shift in the demographic composition and characteristics of the study sample, making what was once a representative sample less representative of the broader population of interest.

Strategies to minimise attrition bias include:

- ✓ Provide easy ways for participants to update their contact details.
- ✓ Use additional and personal reminders via multiple communication channels to motivate those who are not responding to engage with the study, particularly in later waves of data collection.
- ✓ Offer financial or other incentives for study participants to remain in the study until its completion.
- ✓ Critically reflect on how people dropping out of the study may affect its results. This includes taking this into account when you are assessing evidence and explicitly reflecting on this in the evaluation report.

<sup>19</sup> A brief overview of sampling techniques can be found at Qualtrics, 'Your Ultimate Guide to Sampling Methods and Best Practices', accessed 16 January 2023.



**Measurement bias** occurs when the types of information a study collects paint an uneven or incomplete picture of what the study is trying to understand. Integration, for example, is complex, multi-faceted, and a long-term process, and using limited measures of it may lead policymakers to overlook important policy impacts on integration outcomes.

Strategies to avoid measurement bias include:

- ✓ Use holistic measures of migrant integration. For example, when a policy aims to promote language acquisition, it can be helpful to go beyond a standardised assessment of their proficiency level to also measure the impact of the policy on social inclusion, labour market integration, well-being, sense of belonging, health, and other integration outcome that may be related.
- ✓ Use longitudinal studies to capture longer-term integration outcomes whenever possible. When not possible, be aware that a policy's full impact on integration outcomes may take a longer time to become visible.
- ✓ Combine different measures and research methods (including both qualitative and quantitative) to create a more complete picture of migrant integration outcomes.

**Cultural measurement invariance** is a specific form of measurement bias that occurs when the same data collection tools are used with culturally different groups. A question may mean different things to different people. For example, in Western cultures health is often conceived of as a dichotomy between mental and physical health, while some other cultures have a more holistic concept of health. Questions that distinguish between mental and physical health may therefore not collect comparable information across cultural groups because respondents' answers will be shaped by their culture's underlying understanding of health. Given the high level of diversity within migrant and refugee populations, policymakers and evaluators should be aware of how different concepts and questions will be understood by the different groups they work with.

Strategies to overcome cultural measurement invariance include:

- ✓ Use survey questions that have already been validated to reliably measure a specific construct across different groups. Those available in Ethnig's [Ethnic and Migrant Minority Survey Question Data Bank](#) are one example.
- ✓ Use pilot tests to try out draft survey questions. This could include using cognitive interviews, a technique for identifying problems in surveys by pretesting them with a small group of respondents and asking the respondents to verbally describe how they go through the process of interpreting and answering the questions.
- ✓ Test for cultural measurement invariance after collecting the data by using multigroup confirmatory factor analysis. This statistical method allows researchers to assess whether the survey questions (also called items) have equally measured an underlying concept that is not directly measurable, such as mental health, across cultural groups. For example, this type of analysis may determine that one out of ten questions that was used to measure mental health may not measure it accurately among some cultural groups, creating bias in the overall measurement/score on a depression assessment and making

the results incomparable across cultural groups. Recognising this can help evaluators and policymakers understand what the data can—and cannot—tell them.

**Publication bias** is very common and occurs as a result of biases within publication processes that affect what evidence is published. A well-known pattern is that studies that prove that a policy is effective are more likely to be published than studies that show that a policy is *not* effective. This is problematic because it can lead to a body of evidence on an issue that misrepresents reality and the true spectrum of evidence available. It is thus crucial that evidence about what does not work is also published and made available to policymakers and practitioners.

Strategies to address publication bias include:

- ✓ Publish all evaluations, including those that show a policy or intervention has no or a negative impact.
- ✓ When searching for evidence, be aware of this type of bias and try to find conference reports and other reliable sources of evaluation results that may not be formally published and that may discuss a wider range of ineffective as well as effective policies.



*Want to learn more* about overcoming methodological challenges when doing research among migrant populations? Check out:

- ★ The IMISCOE book [Surveying Ethnic Minorities and Immigrant Populations: Methodological Challenges and Research Strategies](#) provides strategies to design high-quality surveys, improve the representativeness of samples, and address many other topics. It also provides examples from surveys that target migrants and ethnic minorities.

### 3.6 Commissioning policy evaluations

When an organisation does not have the in-house capacity to conduct policy evaluations, these are often done by external evaluators on commission. The following tips can help you commission policy evaluations efficiently.

- ✓ Keep a pool of evaluators on file.
- ✓ Encourage potential evaluators to form consortia so all of the skills required for an evaluation are covered in one bid.
- ✓ Organise ‘supplier days’ where potential evaluators can learn more about your intervention and the requirements you have for the evaluation.
- ✓ Create clear assessment criteria to be used when selecting which evaluators will win the bid.
- ✓ Use a two-step assessment process:
  - 1 Assess bids against minimum quality criteria that have been established prior to launching the bid (your must-haves).
  - 2 Assess all bids that pass these minimum criteria based on their quality in proportion to the costs. Decisions at this stage should be based on the goals of the evaluation and the available resources.
- ✓ Plan sufficient time (at least multiple months) for the commissioning process, given it involves multiple steps—from gaining procurement sign-off and providing sufficient time for evaluators to put in a bid, to allowing time for careful assessment and the awarding of the contract.
- ✓ Consult departmental experts on which commissioning route fits best with your evaluation plan.
- ✓ Include information in the request for tenders (RFT, also known as a request for proposals, or RFP) or terms of reference (TOR) about the policy or programme’s logic and about key internal and external stakeholders and their anticipated involvement in the evaluation.



**Tip:** You can use the free [BetterEvaluation GeneraTOR](#) to easily create the terms of reference (TOR) or request for proposals (RFP) for evaluation projects.



**Want to learn more** about commissioning evaluations? Check out:

- ★ The chapter [Commission the Evaluation Project](#) in the Evaluation Toolkit published by the Government of New South Wales in Australia provides a concise overview of best practices on how to develop a request for tenders, handle procurement, and choose the right evaluators, plus tips on contracting.

### **3.7 Further reading and resources**

#### ***Resources on picking an evaluation method and conducting an evaluation:***

- ★ [The Green Book](#), published by the UK government, provides policymakers with guidance on the design and use of monitoring and evaluation.
- ★ [The Magenta Book](#), also from the UK government, includes an overview of the stages of evaluation as well as guidance on how to choose the right evaluation method (see Chapter 3).
- ★ The European Evaluation Society provides lists of [Guidelines and Handbooks](#) and [Tools and Checklists](#) that cover a wide range of evaluation information and practical resources in different languages.
- ★ [World Bank Group Evaluation Principles](#) is a handbook that illustrates the steps to follow when selecting, planning, and conducting evaluations.
- ★ The OECD report [Improving Governance with Policy Evaluation: Lessons from Country Experiences](#) provides an overview of policy evaluation approaches in and practical examples from OECD countries.
- ★ The SPRING Consortium's [Evaluation Tool](#) allows policymakers to quickly self-assess the quality of migrant integration practices and to identify both strengths and areas for improvement.

#### ***Resources on platforms promoting evaluation:***

- ★ [BetterEvaluation](#), part of the Global Evaluation Initiative, aims to support the creation, sharing, and use of knowledge about how to plan, manage, conduct, and use evaluation. Their work covers a knowledge platform, evaluation capacity-strengthening activities, and research and innovation projects.
- ★ The [International Initiative for Impact Evaluation \(3IE\)](#) offers impact evaluation services as well as examples of evidence impact summaries.
- ★ [Knowledge for Policy](#), the European Commission's platform for evidence-based policymaking, provides support to policymakers and researchers working on evidence-informed policy.
- ★ The European Commission's Directorate-General for Regional and Urban Policy's [Impact Evaluation Centre](#) offers policymakers information on policy evaluation, support, and training opportunities.

© 2023 Migration Policy Institute Europe and SPRING Project.  
All Rights Reserved.

Layout: Liz Hall  
Cover photo: iStock.com/Valentyna Ivashchenko

No part of this publication may be reproduced or transmitted in any form by any means, electronic or mechanical, or included any information storage and retrieval system without permission from Migration Policy Institute Europe. A full-text PDF of this document is available for free download from [www.mpieurope.org](http://www.mpieurope.org).

Information for reproducing excerpts from this publication can be found at [www.migrationpolicy.org/about/copyright-policy](http://www.migrationpolicy.org/about/copyright-policy). Inquiries can also be directed to [communications@migrationpolicy.org](mailto:communications@migrationpolicy.org).

Suggested citation: Slootjes, Jasmijn and Maria Belen Zanzuchi. 2023. *Toolkit for Evidence-Informed Policymaking in Migrant Integration*. Brussels: Migration Policy Institute Europe and SPRING Project.