Co-funded by
the European Union

PROGRAMME
EDUCATION
2021-2027

Strengthening Evaluation Practice

Supporting Evaluation across the Thematic Areas of Bulgaria's Programme Education

OECD

Foreword

This report is an output under the OECD's agreement on multi-year support for the implementation of the Evaluation Plan of Programme "Education" 2021-2027 with the Ministry of Education and Science of the Republic of Bulgaria and the Executive Agency "Programme Education". Programme "Education" 2021- 2027 is financed by the European Social Fund Plus and Bulgaria's national budget.

This report was prepared by the OECD Secretariat's Directorate for Education and Skills between May 2025 and April 2026. Anett John (external consultant) was lead author and Cassie Hague led the development of the report and contributed as a co-author to some sections. Zhasmin Avetisyan assisted through research and coordination throughout the report's development. Giles Boden-Tebbutt provided administrative support and Eda Cabbar co-ordinated the overall production of the report. Christa Rawkins leads the OECD's evaluation support to Bulgaria.

Nora Révai from the OECD's Evidence Web of Education project, and external evaluation expert Professor Manuel Souto-Otero from the University of Bristol provided valuable feedback at critical stages of the drafting process. The authors thank both for their invaluable insights and comments. The evaluation team benefitted from the oversight of Elizabeth Fordham, Deputy Head, and Paulo Santiago, Head of the Policy Advice and Implementation Division within the Directorate for Education and Skills at the OECD, under the leadership of Andreas Schleicher, Director for Education and Skills. The authors are grateful to them for their input and for the additional oversight provided by the OECD's Education Policy Committee.

The authors wish to thank Bulgaria's Executive Agency Programme "Education" for their assistance in coordinating data collection and facilitating productive discussions with stakeholders. Their view of evaluation as a learning tool was a driving force behind the report and its areas of focus. Their careful review of the draft versions, and that of the Ministry of Education and Science, are also gratefully acknowledged.

Finally, the authors extend sincere gratitude to all individuals in Bulgaria, as well as to the Managing Authorities of the European Social Fund Plus in Ireland and Poland, who participated in the data collection process. Their input and insights were essential in shaping the findings and recommendations of this report.

Table of contents

Foreword 3

Abbreviations and acronyms 6

Executive Summary 7

1 Overview: purpose, scope and main conclusions 9
1.1 Background: Programme “Education” 2021-2027 10
1.2. Purpose and scope of this report 11
1.3. Data collection methods 15
1.4. Main conclusions 17
References 18
Notes 19

2 Strengthening theories of change 20
2.1. Building a shared understanding of theories of change 20
2.2. Current status and challenges for theories of change in Programme “Education” 23
2.3. Recommendations to strengthen the use of theories of change in Programme “Education” 2021-2027 26
References 29

3 Supporting evaluation thinking and planning among beneficiaries 31
3.1. Current status and challenges for evaluation thinking among beneficiaries 31
3.2. Recommendations for support measures for beneficiaries 33
References 45

4 Strengthening system-level data use 48
4.1. Current status and challenges of system-level data use 48
4.2. Designing Programme “Education” indicators with a data access plan from the start 49
4.3. Anticipating challenges: individual-level data and data protection 50
4.4. Creating a simple, secure route to link and use data across institutions 53
References 54

5 Considering experimental and quasi-experimental evaluation options in procedure design 56
5.1. Enabling conditions and suitable contexts for different quasi-experimental or experimental designs 56

5.2. When might experimental and quasi-experimental methods be feasible in Programme "Education" or in future programming? 58
5.3. Why are advance planning and data access essential? 59
5.4. Practical implications for the Executive Agency Programme "Education" 60
References 61

6 Applying principles of evaluation practice to Programme "Education" 62
6.1. Theory of change: Modernisation of Higher Education Institutions 2.0 62
6.2. Choice of indicators 69
6.3. Data access concerns 70
6.4. How experimental and quasi-experimental designs could be applied to Programme "Education" 74
References 78

Annex A. Indicators 80

FIGURES

Figure 6.1. Procedure-level theory of change for *Modernisation of Higher Education Institutions 2.0* 64
Figure 6.2. Beneficiary-level theory of change for the Modernisation of Higher Education Institutions 2.0 procedure 68
Figure 6.3. Evaluating general support under *Success for You*: A regression discontinuity design 78

TABLES

Table 1. Output indicators (by activity cluster) 80
Table 2. Intermediate outcome indicators 81
Table 3. Final outcome indicators 82
Table 4. Indicators framework for a work package on digital and flexible learning 83

BOXES

Box 2.1. Using indicators to make inferences about the causal impact of a programme 28
Box 3.1. Monitoring and evaluation of European Social Fund Plus projects in Ireland 40
Box 3.2. Financing Not Linked to Costs: Simple illustration examples 43
Box 3.3. Results-based finance: Getting incentives right and managing risks 44
Box 4.1. Monitoring and evaluation of European Social Fund Plus projects in Poland 51

Abbreviations and acronyms

ELA	Poland's Graduate Tracking System
ESF	European Social Fund
ESF+	European Social Fund Plus (previously ESF)
EU	European Union
FNLC	Financing not Linked to Costs
GDPR	General Data Protection Regulation
HEI	Higher education institution
MEL	Monitoring, evaluation and learning
MOOC	Massive Open Online Courses
NSI	National Statistical Institute
PDIA	Problem-Driven Iterative Adaptation
SCO	Simplified Cost Option
STEM	Science, technology, engineering and mathematics
ToC	Theory of change
VET	Vocational Education and Training
ZUS	Poland's Social Insurance Institution

Executive Summary

This report was developed as part of a multi-year cooperation between Bulgaria's Executive Agency Programme "Education", the Ministry of Education and Science and the OECD. Its purpose is to help strengthen evaluation thinking and practice across Programme "Education" 2021-2027, so that evidence can better inform implementation and support long-term impact. The report takes a formative, forward-looking approach and builds on the monitoring and operational systems already in place. It explores how clearer theories of change, more outcome-focused indicators, fuller use of administrative data and more supportive monitoring and learning processes could further help actors across the system understand how and why change is expected to occur. Rather than prescribing a single model, it offers a menu of measures that can be introduced gradually, adapted to specific constraints and refined over time.

The analysis draws on document review, interviews and international examples. It is based on three core challenges. First, while Programme "Education" has clear objectives and a well-established monitoring system, the hypothesised causal links between activities, mechanisms and outcomes sometimes remain implicit. Making these links more explicit can strengthen a shared understanding and support more strategic use of evidence. Second, there is scope to strengthen learning during implementation by creating more space for beneficiaries to reflect on what is contributing to change, for whom, and under what conditions. Third, although Bulgaria has rich administrative data, access and linkage limitations can make it difficult to track outcomes for full cohorts or to construct suitable comparison groups. Addressing these constraints could expand the system's capacity to understand results.

The report first explains how theories of change can help clarify the intended pathways from activities to outcomes. It describes the core components of a theory of change and notes that Programme "Education" already contains many relevant elements, though often not linked together as an explicit causal chain. It then proposes ways to embed simple, practical theories of change at procedure and beneficiary levels.

The analysis then considers how beneficiaries can be supported to integrate evaluation thinking into their everyday work. It recognises the extensive reporting obligations already in place and suggests options to make existing reporting and monitoring processes more oriented towards learning. Examples include application templates based on theories of change, clearer target-setting routines, simplified reporting combined with reflective elements, internal monitoring habits and peer-learning structures.

The discussion then turns to Bulgaria's broader education data landscape. It explains why access to administrative data is essential for tracking whole cohorts to provide a more credible sense of impact and reduce the confounding effects of participant selection. It suggests designing indicators together with explicit data access plans and creating a clear, secure pathway for linking education, social security and labour-market datasets.

A further analysis outlines when more rigorous causal evaluation design may be feasible within Programme "Education". It emphasises that these methods are relevant only for some types of activities, but can add value where rollout, capacity constraints or eligibility rules create natural opportunities to build valid comparison groups. It also discusses the enabling conditions, ethical considerations and data requirements for designs such as random assignment, phase-ins and regression discontinuities.

Finally, the report provides concrete illustrations of how recommendations from earlier chapters can be operationalised within Programme “Education”. It presents example theories of change and indicators for the project, *Modernisation of Higher Education Institutions 2.0*, and illustrates how system-level and beneficiary-level data might be used. It also shows how simple experimental or quasi-experimental approaches could be applied to selected activities where conditions allow, taking *Success for You* - an inclusive education intervention in schools - as an example.

Key findings

The report identifies several interconnected opportunities that can build on existing strengths to enhance evaluation practice in ways that support high-quality implementation and long-term impact across Programme “Education”.

1. **Making expected causal pathways explicit through simple, shared theories of change:** Explicit and clear theories of change at procedure and project levels could help articulate the targeted mechanisms of change, make underlying assumptions visible, and link activities to the outcomes they seek to influence. They could support more coherent planning, more focused indicator selection and earlier recognition of where additional support may be helpful.
2. **Embedding evaluation thinking in everyday implementation to strengthen learning and adaptability:** Supporting beneficiaries to reflect on how their activities are expected to lead to change, and to check these assumptions against routine monitoring data, could strengthen learning during implementation. This could help institutions identify bottlenecks, unintended effects and differences in what works across contexts. Supports such as workshops, peer exchange, coaching and internal monitoring may help institutions use their own data to refine activities and respond to emerging needs.
3. **Developing indicators that capture mechanisms and outcomes based on the theory of change:** Indicators could usefully be derived from the theory of change and used to track outputs, intermediate outcomes and final outcomes. Where possible, outcomes could be measured at cohort level rather than only for participants, to provide a more robust sense of Programme effects and reduce confounding from participant selection.
4. **Designing indicators with feasible data access in mind can make monitoring more informative and efficient:** Including a simple “data access plan” when defining indicators – for example, specifying which data are needed, who holds them and how they can be accessed securely - could help ensure that indicators remain measurable over time, reduce ad hoc data requests and streamline reporting for beneficiaries.
5. **Strengthening system-level data access and infrastructure can expand evaluation possibilities:** Bulgaria’s existing administrative systems already contain much of the information needed to track learning outcomes, progression and labour-market trajectories. Establishing a clear and secure route for linking datasets while fully respecting data-protection requirements could support more robust evaluation, allow for tracking whole reference groups, and reduce the need for bespoke data collection.
6. **Designing selected procedures with evaluation in mind can create opportunities for feasible quasi-experimental and experimental studies:** Where activities involve phased rollout, capacity constraints or well-defined eligibility thresholds, these features can support experimental or quasi-experimental evaluation without disrupting delivery. The Executive Agency Programme “Education” and the Ministry of Education and Science could consider this in the design of procedures and projects to help make these methods ethically and operationally feasible for both internal and independent evaluators.

1 Overview: purpose, scope and main conclusions

This report aims to support the Executive Agency Programme “Education” (hereafter, Executive Agency) and the Ministry of Education and Science of the Republic of Bulgaria (hereafter, the Ministry) to strengthen evaluation practice for Programme “Education” 2021-2027 and future programming. It was prepared under the Agreement on multi-year support for the implementation of the Evaluation Plan of Programme “Education” 2021-2027 between the Executive Agency, the Ministry, and the OECD.

The report focuses on practical ways to make evidence and evaluation more actionable for programme design and delivery in Bulgaria. It examines how clearer and more consistently used theories of change can help articulate the problems that procedures aim to address, clarify the mechanisms through which activities are expected to work, identify key assumptions and risks, and translate these into a small set of reasonable and measurable indicators.

A stronger use of theories of change is especially timely in Bulgaria, where Programme “Education” finances a broad set of reforms across diverse thematic areas and implementation is delivered through many beneficiaries. In such settings, explicit theories of change can support a shared understanding of objectives and pathways to results, strengthen monitoring and learning during implementation, and improve the credibility and usefulness of evaluation findings.

The report therefore examines three interconnected areas: (i) How to strengthen the integration of theories of change and indicators into programme- and project-level planning and delivery; (ii) How those that manage projects can be supported to use their monitoring data for reflection and learning, so that implementation generates more useful evidence; and (iii) How access to and use of system-level data can be improved, contributing to a more coherent and comparable evidence base for planning, monitoring and evaluation.

The report is structured as follows:

- Chapter 1 sets out the background to the report, the questions it addresses, and the data collection methods it uses. It also provides a summary of the main conclusions.
- Chapter 2 focuses on theories of change.
- Chapter 3 considers how beneficiaries could be further supported to integrate reflection and evaluation-oriented practices into implementation so that delivery generates usable evidence on what is working and why.
- Chapter 4 looks at access to data within the Bulgarian education system.
- Chapter 5 explores how choices made during the design of procedures can affect the feasibility of different kinds of impact evaluations.
- Chapter 6 applies the principles covered in earlier chapters to specific areas of Programme “Education” to illustrate what they might look like in relation to concrete procedures and activities.

1.1 Background: Programme “Education” 2021-2027

Programme “Education” is Bulgaria’s main programme in education funded by the European Union (EU) for the 2021-2027 period. It is co-financed by the European Social Fund Plus (ESF+) and national resources, with a total budget of almost EUR 965 million[1]. Programme “Education” succeeds the Operational Programme “Science and Education for Smart Growth” 2014-2020 and is designed to support system reforms that improve access, inclusiveness, quality, and labour market relevance across the education sector.

Programme “Education” is closely linked to Bulgaria’s long-term development agenda. It is one of the main tools for implementing Priority 1 “Education and skills” of the National Development Programme BULGARIA 2030. It also supports national strategic frameworks for pre-school, school, vocational and higher education, including the Strategic Framework for the Development of Education, Training and Learning (2021-2030) and the Strategy for the Development of Higher Education (2021-2030). In this way, Programme “Education” contributes to Bulgaria’s broader objectives to raise economic growth, address demographic decline, and reduce inequalities.

Programme “Education” is organised around three priorities: (1) Inclusive education and educational integration, including for marginalised groups such as Roma and other vulnerable groups; (2) Modernisation and quality of education; and (3) Relevance of education to the labour market (EAPE, 2024[1]). These are translated into five broad policy objectives:

1. Improving equal access to inclusive, high-quality education and raising participation and completion, especially for vulnerable and marginalised learners;
2. Strengthening digital education and the digital competence of teachers, students and parents;
3. Introducing innovative curricula and learning content and supporting the professional learning of educators;
4. Improving the labour market relevance of vocational and higher education, including through dual learning;
5. Modernising and digitalising higher education and fostering entrepreneurial skills among students.

The scale of Programme “Education” is substantial. Programme documents indicate that planned operations are expected to reach more than 760 000 children and students (EAPE, 2024[1]). Around 300 000 children, students and parents from marginalised groups, including Roma, are expected to receive targeted support. Measures to improve digital skills are expected to benefit nearly 370 000 children and students. In vocational education and training, modernisation measures are expected to reach almost the entire public sector, with about 380 schools participating out of 391 public vocational schools as of 2023-2024 (NSI, 2024[2]) and roughly 14 000 students will receive support in their transition to the labour market (EAPE, 2024[1]). The budget for Programme “Education” is more than double that of the previous programming period, reflecting both the ambition of the reforms and the central role of education in Bulgaria’s development strategy.

The ESF+ is the EU’s main financial instrument for investing in people. With an EU wide budget of around EUR 88 billion for 2021-2027, ESF+ supports Member States in reducing unemployment, improving the quality and equity of education and training, and strengthening social inclusion. Funds are invested in national and regional programmes such as Programme “Education”. ESF+ is aligned with the European Pillar of Social Rights and promotes horizontal principles such as respect for fundamental rights and compliance with the Charter of Fundamental Rights of the EU, gender equality, and non-discrimination. Through Programme “Education”, Bulgaria uses ESF+ resources to advance these objectives in the education sector and to contribute to the EU’s 2030 social targets.

1.2. Purpose and scope of this report

Given the level of ambition and investment articulated through Programme “Education”, robust monitoring, evaluation and learning (MEL) is essential. Bulgaria has put in place an Evaluation Plan for Programme “Education” 2021-2027 (EAPE, 2023[3]), as required by EU Regulations, and has commissioned this report as part of its ongoing work to strengthen the evidence base for programme delivery and reporting.

The Executive Agency, which is the Managing Authority for Programme “Education”, and its beneficiaries routinely collect detailed implementation data on the operation and project levels, including information about implemented activities, achieved outputs and results, financial expenditures and project participants. This ensures that Programme “Education” is fully aligned with ESF+ regulations and requirements on indicators, reporting and audit. The purpose of this report is therefore not to suggest that more monitoring and evaluation is introduced to the system. Instead, it builds on the existing architecture to support a cycle of reviewing current practice and informing continuous improvement in how the quality and impact of funded interventions are measured and improved.

The primary audience of this report is the Executive Agency. It is also intended to be of relevance to policymakers and practitioners across the Bulgarian education system as well as partners and stakeholders involved in the design, implementation and evaluation of this and future programmes. The report can also be of interest to those invested in advancing evaluation thinking and practice more widely, both inside Bulgaria and internationally.

Defining evaluation practice

In this report, evaluation practice refers to the systematic and objective assessment of education policies, programmes and projects, including their design, implementation and results (OECD, 2023[4]). Effective MEL helps ensure that resources reach the intended target groups, that funded activities are implemented at high quality, and that they contribute to better and more equitable outcomes. It strengthens understanding of how target groups experience programmes, provides early signals when measures are ineffective, supports learning about what works in different contexts, and underpins transparent reporting to national stakeholders and to the European Commission.

Educational evaluation remains a relatively young and evolving field, and many countries are still developing approaches that combine accountability with learning and practical use in policy design and implementation. The suggestions in this report aim to further support the development of evaluation practice in Bulgaria and encourage system-wide reflection on how MEL processes operate and could be improved to advance the programme’s goal of achieving lasting benefits for learners and other stakeholders involved in the Bulgarian education system.

The importance of evaluation thinking

All policies and activities in Programme “Education” ultimately seek to improve young people’s lives. In education this means, for example, helping more children attend pre-school, reducing early school leaving, improving learning outcomes, strengthening the inclusion of those from marginalised communities, such as Roma, and supporting the employability of graduates. The current performance framework is designed according to the requirements of Regulation 2021/1060 to ensure effective monitoring, reporting and evaluation of programme performance. It provides important information on how funds are spent and on the scale of activities, such as the number of teachers trained, or students reached. However, meeting these process targets is not, on its own, a guarantee that the deeper educational goals are being achieved.

This is where evaluation thinking comes in. Evaluation thinking complements existing monitoring by encouraging all actors to ask a set of simple but powerful questions from the very beginning of a project or operation: What specific changes do we want to see for learners, teachers or institutions? Through which

mechanisms do we expect these changes to occur? Under which conditions is the intervention likely to work, and for whom? What information will we need, and when, to see whether we are on the right path and to correct course if needed? Establishing the importance of these questions early helps to direct attention beyond the direct inputs and outputs of activities towards outcomes and the mechanisms that can produce them. Some existing results indicators already point in this direction, such as the number of students with improved educational results one year after participation in an operation. The report explores how such outcome-oriented elements can be further strengthened.

An important aspect of evaluation thinking is recognising the underlying purposes of evaluation itself. Evaluation is not intended to be punitive, focussed only on whether beneficiaries comply with funding requirements, or limited to accountability alone. At its core, evaluation is a knowledge-building practice that helps beneficiaries and policymakers understand what is working, for whom, and under what conditions (Pawson and Tilley, 1997[5]; OECD, 2023[4]). Seeing evaluation in this way can help reduce resistance and uncertainty around evaluation amongst those that manage projects. This view of evaluation also supports ongoing reflection and adjustment of implementation and MEL approaches when activities do not unfold as expected. Evaluation thinking therefore serves as a basis for improved evaluation practice, which in turn can contribute to stronger impact.

A central part of evaluation thinking is to make the theory of change explicit. This means describing, in simple terms, how the planned activities are expected to lead to outputs, intermediate changes and final outcomes. It also means identifying key assumptions and risks and agreeing which indicators will be used to track each step in the chain. When a theory of change is clear and shared, it becomes easier to spot bottlenecks along the path from activities to outcomes (e.g. a lack of teacher capacity, or low uptake among the intended target group), to introduce real-time feedback loops and quality assurance mechanisms, and to adjust implementation before final outcomes are measured. A well-articulated theory of change thus strengthens programme design and implementation and provides a clearer basis for any later evaluations.

This report also shows how evaluation practice can be embedded more systematically in programme design and day-to-day delivery to generate usable evidence during implementation and provide a stronger basis for later independent evaluation. When evaluation is added as a standalone component at a later stage, after the initial roll-out of a programme, important opportunities are easily missed: data that would have been useful may not be collected; no comparison group is created; and delivery is not documented in ways that allow meaningful learning. By contrast, integrating evaluation considerations at the conception stage makes it easier to design operations and procedures that both maximise impact and generate usable evidence.

Scope of this report

The Executive Agency has established an agreement with the OECD to support a multi-year, learning-focussed, evaluation of Programme “Education”. The OECD is producing a series of interim, final and ad hoc evaluation reports to explore what works, for whom, in what contexts, and how. The aim is to provide recommendations that are grounded in a careful analysis of the contextual factors that influence outcomes and that can be used to enhance the implementation of Programme “Education”.

Previous work by the World Bank (2021[6]) laid out a comprehensive Monitoring and Evaluation methodology for Bulgaria’s education sector, aligned with national and sectoral strategies. Since then, efforts have been made to strengthen the institutional landscape for evaluation, most recently with the establishment of the Institute of Education in 2024. The Institute brings together several existing bodies – the Centre for Assessment in Preschool and School Education, the Centre for Information Provision of Education, and the Centre for Training Companies – to provide a consolidated structure for research, methodological guidance, and participation in national and international evaluation activities. In cooperation with the World Bank, the Executive Agency has also conducted a comprehensive review of its indicator framework (EAPE, 2022[7]). In addition, based on recommendations from an internal assessment

of projects with the Ministry as a direct beneficiary conducted in 2022, it introduced monitoring questionnaires for most procedures. These developments signal a clear commitment to strengthening the evidence base for programme management and reporting.

Building on these foundations, the current report focuses on translating the principles of evaluation thinking into practical, context-sensitive recommendations for current and future programming. It aims to support the Executive Agency in identifying feasible steps to strengthen evaluation practice in programme management and among beneficiaries, without overstretching existing systems or capacities. An important focus is on helping projects articulate clearer theories of change, so that intended outcomes – such as improved learning, inclusion, or employability – and the mechanisms through which they are expected to be achieved are clearly defined from the outset and can guide implementation. These theories of change can help inform decisions about what to measure and build shared expectations around what constitutes success.

The report explores how evaluation systems and processes can be strengthened incrementally to better reflect those intended outcomes. This includes identifying practical ways to move beyond participation indicators and begin capturing more meaningful results that can help determine outcomes or impact (e.g. behavioural changes, improvements in school practice, progress in reaching underserved groups). The report identifies opportunities to support learning and usability while remaining realistic about institutional constraints. A key aim is to recommend improvements that are not only feasible, but that also support ownership and engagement among stakeholders. Theories of change can increase ownership by placing stakeholders' views at the centre of the model of change and making their assumptions explicit in evaluation questions, indicators and methods. This in turn can support more participatory forms of evaluation and increase the likelihood that findings are accepted and used by those responsible for education change.

Finally, the report provides concrete illustrations of how the suggested evaluation practices could be applied to different types of projects and beneficiaries within Programme "Education". To do this, it draws on three procedures, covering diverse thematic areas and end beneficiaries, as worked examples *Modernisation of Higher Education Institutions 2.0*, which focuses on strengthening competence-based models in higher education; *Support to Centres of Excellence in VET*; and, *General and Additional Support for Personal Development in School Education* within the inclusive education theme. While the latter has been running for more than two years, the other two are currently being launched.

These examples are used to illustrate how selected recommendations from the report could be translated into practice for specific activities, for instance by clarifying theories of change, refining indicators, planning simple feedback loops, or considering options for data use and impact evaluation, where feasible. The aim is to show what the proposed changes might look like in practice and to support the Executive Agency and beneficiaries to adapt these ideas to their own operations.

Ultimately, the report aims to help strengthen applied and reflective evaluation practice in educational programming in Bulgaria with the goal of enabling both the Executive Agency and beneficiaries to use evidence effectively to support learning, accountability, and programme improvement.

Questions to guide the analysis

As this report does not evaluate specific projects or procedures under Programme "Education", the six evaluation criteria agreed by the OECD and the Executive Agency (relevance, coherence, effectiveness, efficiency, EU added value and sustainability) are not applicable. Instead, the focus is on identifying how evaluation thinking, monitoring frameworks and grant procedures can be strengthened across the programme to support more systematic and robust evaluation practice in the future.

The report will be guided by questions in four key areas.

A: Strengthening theories of change to clarify outcomes and causal pathways

How can the Executive Agency, beneficiaries and other stakeholders develop, refine, and use theories of change that genuinely support project design, measurement of outcomes, and learning?

- In what ways can theories of change be used as practical tools throughout the project cycle to guide design and planning, inform adaptation, and support meaningful evaluation?
- How can theories of change make intermediate outcomes and mechanisms more explicit – linking high level programme goals to the factors that plausibly move them (e.g. changes in teaching practice, attendance, or learner engagement)?
- How can beneficiaries and other stakeholders be supported to develop theories of change that consider the full range of possible effects, including both intended outcomes and unintended consequences, such as counterproductive incentives or strategic behaviour among participants?

B: Supporting evaluation thinking and planning among beneficiaries.

How can evaluation become a tool that genuinely helps beneficiaries understand and improve the impact of their work?

- How can reporting requirements be designed in a way that supports learning, ownership, and honest reflection, and encourages a culture where difficulties can be surfaced early without fear of failure?
- What kinds of support (e.g. peer learning, coaching, worked examples) are most effective in helping beneficiaries reflect on what works, for whom, and under what conditions?
- What incentives or feedback loops would encourage beneficiaries to actively engage with their own evaluation data, while reducing disincentives to report implementation challenges or mixed results honestly?
- What capacities (i.e. technical, organisational, or relational) do beneficiaries need to plan, interpret, and act on evaluation results and where are the key gaps?

C (where relevant): Strengthening system-level evaluation processes and data use

While not the primary focus of the report, system-level evaluation processes and data use may be addressed where they present barriers to effective evaluation by programme beneficiaries or to the use of evaluation findings. In such cases, the report will consider how these processes and data systems could better support meaningful insights and contribute to evidence-informed decision-making.

- What are the main limitations in the current data landscape that make it difficult to carry out effective evaluation in education?
- How can the existing data infrastructure, information and communication technology systems and legal framework be used to better monitor the impact of the activities and to potentially condition beneficiary incentives?
- What types of data – beyond existing administrative indicators – would help capture the broader outcomes that Programme “Education” aims to achieve?
- What capacities and institutional arrangements are needed to ensure that available data can be aggregated, analysed and used effectively for monitoring and evaluation?

D (for selected activities): Designing procedures and activities with experimental and quasi-experimental methods in mind

In a limited number of cases, it may be feasible and appropriate to design activities in ways that allow future experimental or quasi-experimental evaluation, without disrupting service delivery.

- In which types of operations (e.g. where support is phased in over time, where demand exceeds available places, or where eligibility uses sharp cut-off rules) could simple experimental or quasi-experimental designs be considered?
- How can eligibility rules, roll-out schedules or oversubscription procedures be specified so that they remain fair and transparent while also creating plausible comparison groups for later impact evaluation?
- What minimal design and data elements (e.g. stable identifiers and agreed outcome measures) would need to be in place to make such evaluations feasible and ethical at reasonable cost?

1.3. Data collection methods

This report is primarily formative in nature, with a focus on future improvements. It does not assess the effectiveness of individual operations under Programme "Education" but examines how evaluation practice and monitoring frameworks could be strengthened across the programme. To a substantial extent, the work is thus theory informed: it focuses on how design features (e.g. theories of change, indicator frameworks, reporting templates, grant rules, data access arrangements) shape the way actors think about impact, monitor results, and learn from experience.

To identify current strengths as well as areas for improvement, the OECD evaluation team used purposive sampling, deliberately selecting individuals and materials most directly connected to the design, management and use of Programme "Education" and comparable ESF+ programmes in other countries. Data collection was iterative: early insights from documents and interviews helped refine the focus of subsequent interviews, the selection of international examples, and the emerging recommendations.

Data collection took place between May and September 2025; the following methods were used:

- **Document review of Programme "Education"** helped map the programme's objectives, intervention logic and existing monitoring framework. This included the Programme "Education" document, application and reporting templates; technical guidelines; indicator factsheets; and, where available, documentation from previous procedures including direct awards and competitive calls. These materials were used to understand how theories of change are currently expressed, how indicators are defined and how monitoring and reporting requirements are embedded in grant procedures.
- **Review of national documentation** provided the context for evaluation and data use in Bulgaria. This included key legislative and strategy documents related to education, monitoring and evaluation, and data protection and access (e.g. provisions governing administrative data, interoperability between registries and the role of the National Statistical Institute). These documents were used to identify opportunities and constraints for using system-level data in programme evaluation and for accessing administrative registers for public interest purposes.
- **Review of promising practices in other countries** focussed on how comparable programmes in other EU and OECD countries structure their MEL requirements within grant schemes. The team examined guidance and examples on the use of theories of change in ESF+ programmes, indicator design and target setting, simplified cost options and Financing Not Linked to Costs, and models for secure access to administrative data for evaluation purposes. The aim was not to import entire models as blueprints, but to identify practical MEL design features that could be adapted to the Bulgarian context.
- **Online interviews in Bulgaria** provided insight into how evaluation and monitoring are understood and used in practice. The team conducted seven semi-structured online interviews with two groups of stakeholders: 1) those directly involved in implementing Programme "Education", including programme management teams and beneficiaries of previous activities; and 2) actors from the

broader monitoring and evaluation system, such as national institutes and data-producing agencies that support system-level monitoring, evaluation and evidence use. These interviews explored, among other topics, how reporting requirements function in practice, how indicators and targets are developed, and what support beneficiaries receive when planning and interpreting evaluation activities.

- **Interviews with Managing Authorities of ESF+ programmes in other countries** offered comparative perspectives on reporting and evaluation practice. Online discussions with representatives from Ireland and Poland focused on their approaches to integrating evaluation thinking into calls for proposals and contracts, the use of outcome indicators and, where relevant, results-based financing, and the way administrative data are accessed and used for monitoring and evaluation.
- **Review and verification meetings** were used to validate emerging findings and refine recommendations. The OECD team participated in two online meetings with the Executive Agency, in June 2025 and November 2025, to scope where guidance would be helpful, present preliminary recommendations, check factual accuracy, and collect additional information regarding needs or constraints. Feedback from these exchanges was incorporated into the final analysis.

In interpreting the evidence, the OECD evaluation team also drew on experience from previous related interim evaluations and analytical work on education and evaluation in Bulgaria and other countries. This broader experience informed the identification of changes that are likely to be feasible within existing administrative capacities and regulatory frameworks.

Limitations

This ad hoc report has several limitations that readers should bear in mind when interpreting its analysis and recommendations. Many of these reflect intentional decisions about scope and approach but because they may also be viewed as constraints they are presented and justified here.

First, the report does not assess the effectiveness, efficiency or impact of specific operations or procedures under Programme “Education”. Nor does it evaluate individual beneficiaries. Instead, the scope of the work is deliberately system-focused and formative. It examines how programme design features, such as theories of change, indicator frameworks, reporting templates, grant rules and data access arrangements, shape evaluation practice. As a result, the findings point to areas where monitoring and evaluation could be strengthened in general, rather than providing project-level judgements.

Second, the evidence base is necessarily partial. Interviews were conducted with a small number of stakeholders who either have direct responsibilities for the management of Programme “Education”, are beneficiaries of activities under the programme or hold system-level roles in monitoring and data provision. These stakeholders are well-placed to speak to the design and administration of Programme “Education”, but they cannot be considered representative of all actors involved in implementation. In particular, the report does not incorporate systematic evidence from frontline practitioners or programme participants across the country, and it relies on self-reported descriptions of how procedures operate in practice.

Third, the document review is confined to materials that were available and shared during the research. Programme documentation, indicator factsheets, application and reporting templates, and relevant legal and strategic texts provide a strong basis for understanding formal requirements, but they may not capture less formalised practices that emerge during implementation, nor local adaptations or very recent changes introduced after the period of data collection.

Fourth, international examples are used for illustration rather than as ready-made models. The report draws selectively on guidance and practice from other EU and OECD countries to highlight design options that could inspire improvements in Bulgaria. These examples were not identified through a systematic review of all international experience, and they come from systems with different institutional histories,

legal frameworks and administrative capacities. Any transfer of such practices will require careful adaptation to the Bulgarian context and close attention to national legal requirements, including those on data protection and financial management.

Finally, the recommendations are forward-looking and need to be planned as part of a long-term process of continually improving evaluation practice in Bulgaria. Some recommendations are immediately actionable; others are intended as starting points for discussion, capacity-building and iterative change, not as prescriptions that should be implemented mechanically. Proposals to strengthen theories of change, introduce more outcome-oriented indicators and expand results-based financing will need to be balanced against other constraints, such as the need to keep documentation requirements manageable and to respect the administrative burden on beneficiaries. The practical feasibility, timing, and value of each recommendation will depend on local context, available resources and policy priorities, and should be tested through piloting, feedback from beneficiaries and subsequent evaluations.

1.4. Main conclusions

The report offers a menu or toolbox of measures that the Executive Agency and beneficiaries can draw on to further strengthen evaluation thinking and practice over time. The aim is not to prescribe a single model, but to provide options that can be considered, adapted, tested and refined incrementally as Programme “Education” evolves, as well as in future programmes. The report is therefore designed to inform ongoing reflection, capacity-building and continuous improvement rather than one-off reforms.

The key conclusions of the report, embedded across the remaining chapters, are as follows:

1. **Strengthening the use of theories of change can help ensure a clear and shared understanding of how activities are expected to contribute to change**: Developing more explicit, shared theories of change at procedure and beneficiary or project level would help clarify how particular activities are expected to influence outcomes, through which mechanisms, and what organisational, implementation, logistical or policy conditions need to hold for this to happen. This can also guide the choice of indicators. Providing templates, examples, capacity-building workshops and regular reviews of theories of change could help to make them a practical organising tool that is useful for day-to-day planning and monitoring.
2. **Supporting all actors to reflect on what is working and not working to create change throughout implementation can strengthen learning and make delivery more adaptive**: A stronger focus on evaluation thinking throughout the programme cycle can help beneficiaries and the Executive Agency use data to understand mechanisms, identify bottlenecks and adjust during implementation, rather than relying solely on compliance reporting or external evaluators. Providing practical supports, such as indicator guidance, peer exchanges or coaching, can help beneficiaries make sense of their own monitoring data and understand more about what works in their contexts.
3. **Using theories of change to continue to improve indicators can make progress easier to assess and interpret:** Starting from the theory of change can help identify indicators for outputs, intermediate changes and final outcomes. Where possible, tracking outcomes for wider cohorts, rather than only for participants, can provide a more credible picture of programme effects and reduce confounding from participant selection. Supporting beneficiaries to link these indicators to the intervention logic can help them focus on mechanisms and early signals of change.
4. **Designing indicators with data sources and access in mind from the outset can make monitoring more informative and efficient**: Including a short “data access plan” in indicator descriptions, which clarifies what data are needed, who holds them, and how they can be accessed, can prevent ad hoc requests and ensure outcome indicators are measurable in practice.

5. **System-level data access and infrastructure need strengthening to support more robust evaluation:** Bulgaria has substantial administrative data, but access, linkage and interoperability challenges limit their use for monitoring and evaluation. Improving data governance and enabling secure, lawful access to linked datasets would expand the scope for tracking learner trajectories and conducting rigorous analysis.
6. **Designing some procedures with evaluation in mind can create opportunities for feasible, low-burden experimental or quasi-experimental impact assessments**: Where appropriate, implementation features such as phased rollout, capacity constraints or threshold eligibility criteria can facilitate credible comparisons without disrupting delivery. The Executive Agency and Ministry could consider planning for these options early in the design of programmes and procedures to help make them ethically and operationally feasible, for both internal and independent evaluators.

References

EAPE (2024), *Programme "Education" 2021-2027*, https://sf.mon.bg/?go=page&pageId=364 (accessed on 22 May 2026). [1]

EAPE (2023), *Plan for the evaluation of Programme "Education" 2021-2027*, https://sf.mon.bg/?go=page&pageId=572 (accessed on 22 May 2026). [3]

EAPE (2022), *Methodology for defining the indicator framework*, https://sf.mon.bg/?go=page&pageId=364 (accessed on 22 May 2026). [7]

NSI (2024), *Professional schools by type*, https://dev.nsi.bg/bg/content/3481/%D0%BF%D1%80%D0%BE%D1%84%D0%B5%D1%81%D0%B8%D0%BE%D0%BD%D0%B0%D0%BB%D0%BD%D0%B8-%D1%83%D1%87%D0%B8%D0%BB%D0%B8%D1%89%D0%B0-%D0%BF%D0%BE-%D0%B2%D0%B8%D0%B4 (accessed on 6 April 2026). [2]

OECD (2023), *Glossary of Key Terms in Evaluation and Results-Based Management for Sustainable Development (Second Edition)*, OECD Publishing, Paris, https://doi.org/10.1787/632da462-en-fr-es. [4]

Pawson, R. and N. Tilley (1997), *Realistic Evaluation*, https://uk.sagepub.com/en-gb/eur/realistic-evaluation/book205276 (accessed on 22 May 2026). [5]

World Bank (2021), *Bulgaria: A report detailing a proposal for monitoring system of education policy developments and outcomes*, World Bank Documents & Reports, https://documents1.worldbank.org/curated/en/099336512232138339/pdf/P1713420cc02b203d08e4f03a54765f1907.pdf (accessed on 22 May 2026). [6]

Notes

[1] Programme “Education” is currently undergoing an amendment. In particular, this involves, a transfer of funds to the Food and Basic Material Support Programme and the addition of a new Priority 5 on the civil preparedness of the education system. Following these changes, the Programme’s budget will amount to just over EUR 940 million.

2 Strengthening theories of change

This chapter aims to support a shared understanding of theories of change (ToCs) and how they can inform evaluation thinking and practice at all stages of programming. The chapter discusses why it could be helpful to further embed the use of ToCs in Programme "Education" 2021-2027 and how they can be used in day-to-day management. This can also inform the design of the next programming period. It then looks at the planning and monitoring tools already used in Programme "Education" - including programme documents, guidelines for applicants, indicator factsheets and logical matrices - and considers where elements of ToCs are already present, even if not yet brought together in a single map. Practical examples of ToCs and recommendations on how explicit ToCs could be used in relation to specific programme activities are presented in Chapter 6.

2.1. Building a shared understanding of theories of change

A core theme of this report is the importance of using ToCs in a deliberate and practical way. While their systematic use is not yet widespread in Bulgaria, this is beginning to change, including through recent efforts by the Executive Agency Programme "Education" (hereafter, Executive Agency) to integrate ToCs more explicitly into programme management. The intention of this section is therefore not to introduce a new concept, but to set out how research suggests that ToCs should be defined and used (Weiss, 1995[1]; Rogers, 2014[2]; Mayne, 2017[3]), and to clarify how this report uses the term. A shared understanding of ToCs – including agreement that they bring together key components such as activities, objectives, outputs, outcomes, assumptions and contextual factors – is important not only so that subsequent sections can be interpreted consistently, but also to support more coherent and aligned evaluation practice across education stakeholders in Bulgaria.

In this report, a ToC is understood as an explicit description of the challenge a policy or programme addresses, and how and why the planned activities are expected to lead to the outcomes and impacts they seek to achieve. For the purposes of Programme "Education", the key components of a ToC can be summarised as follows:

- **Inputs and activities** are the resources and actions under direct control of the programme or project team. Examples include funding, staff time, teacher training sessions, coaching visits, the provision of equipment, or the development of new learning materials.
- **Outputs** are the immediate products of these activities. They are usually still under direct control. Examples include the number of teachers who complete a training module, the number of schools that receive equipment, or the number of guidance sessions delivered.
- **Intermediate outcomes** are results that are closely connected to the activities, but already represent changes in behaviours, practices or conditions. They are often instrumental outcomes (i.e. they matter mainly because they are expected to contribute to later goals that are valued in their own right) and therefore provide early signals about whether the intended mechanism is working. For example, after teacher training, a relevant intermediate outcome is whether teaching practices in the classroom have changed. Alternatively, after the provision of digital resources to schools, an intermediate outcome might be the effective use of those resources.

- **Final outcomes** are the shorter- and longer-term changes that an activity ultimately aims to influence, and that are directly relevant to welfare and policy objectives. In education, these might include higher levels of student learning and skills, improved completion rates, better inclusion of learners from marginalised groups, or stronger employment outcomes for graduates. For example, a teacher training programme does not aim only to change teaching practices in isolation. Depending on the setting, its ultimate purpose may be to improve academic achievement, reduce early school leaving, strengthen inclusion, or support smoother transitions into the labour market. Making these long-term goals explicit from the start helps align expectations across different levels of the system.
- **Assumptions** are the conditions that must hold for the links in the ToCs to work as expected, while **risks** are the potential obstacles that could disrupt them. They are typically outside the direct control of the programme, but they can be monitored and, in some cases, influenced. For example, assumptions may include political stability, sufficient staffing or buy-in from school leaders and teachers. Making these assumptions explicit encourages teams to pay attention to them during implementation and to plan mitigating actions for risks where needed.

In practice, ToCs do not need to be complex or technical. A one-page diagram with a short narrative description is often sufficient (see Chapter 6, Figures 6.1 and 6.2), provided that it clearly sets out the main steps, the associated assumptions and the key linkages.

A ToC can be thought of as a map of hypothesised pathways of change. Rather than just listing or describing the components above, a ToC makes the causal links between them explicit – typically in diagram form. In a ToC diagram, arrows show how specific activities are expected to lead to specific outputs, how these outputs feed into intermediate outcomes, and how these, in turn, contribute to final outcomes. Optionally, assumptions can be shown alongside the arrows where they matter most. This "visual web" of connections highlights the mechanisms that are expected to move results and thus reveals the points where bottlenecks are most likely to appear.

A ToC can be specified at different levels of the system. At procedure and project level, the ToC tends to be more general and strategic, describing how broad packages of support (e.g. inclusive education measures, digitalisation, or vocational education and training [VET] reforms) are expected to contribute to national education objectives. At activity level, the ToC can be more concrete, specifying target groups, delivery models and the mechanisms through which the intervention is expected to work. Links can be described in specific terms, such as how a particular coaching model in a school is expected to change classroom practice and student engagement. These different levels should be consistent with each other: an activity-level ToC should fit within the project-level ToC, which in turn should link to the procedure-level.

A ToC can be refined over time and should be reviewed periodically (e.g. annually), or when monitoring or evaluation reveal important deviations, bottlenecks or unintended effects. In most cases, this does not mean changing the ultimate objectives of an intervention but rather refining the understanding of how change is expected to happen in practice. This may involve revisiting assumptions, adjusting activities or sequencing, or recognising mechanisms that were weaker or stronger than expected. For example, if competence-based curriculum reform appears to improve perceived relevance but also increases student workload and lowers completion rates, the ToC may need to be refined to show that positive effects depend on streamlining or removing older content, rather than simply adding new components. Responsibility for updating the ToC should remain with those both managing and implementing the activities. At procedure level, they can inform adjustments to guidance, support and monitoring; at beneficiary level, such reviews can support adaptive implementation.

Why strengthen the use of theories of change?

A central question is, why it is worth investing time in a ToC? For a large programme such as Programme "Education", the main reason is simple: public funds and project resources are limited. An activity with limited impact uses up money, time and attention that could have supported a more effective intervention. A clear ToC is one of the most practical tools available to ensure that scarce resources are used where they are likely to have the greatest effect. It provides a transparent explanation of why particular activities are funded, in a form that others can review and discuss. When the assumptions and links in the ToC are made visible, they can be tested against emerging evidence and adjusted over time so that funding decisions become progressively better aligned with what is known to work. In this sense, a well-specified ToC is not an additional administrative task; it is a practical management tool that helps focus attention on quality and impact. It does this by:

- **Making the goal and the path explicit**. It sets out what will be done, what should change along the way, and what success looks like at the end. When this map is visible, actors at different levels (e.g. Executive Agency, ministries, regional authorities, schools and higher education institutions) can check that they share the same understanding of objectives and of how their own role contributes to them. This can help to reduce misunderstandings and align expectations.
- **Turning big goals into a small set of practical checks during implementation**. By identifying a few key intermediate outcomes, teams can track progress while the project is still running. In a teacher training initiative, for example, the ToC would not stop at "teachers trained". It would also specify expected changes in classroom practice, learner engagement or time on task. These can then become concrete, observable checks during implementation.
- **Helping to judge not only whether an intervention is working, but also how it is working**. Because the ToC spells out steps and mechanisms, it is easier to see where the chain is breaking and to act early. Consider a project that offers remedial teaching after school to improve students' learning outcomes. If the real constraint is irregular attendance, for example because students have caring responsibilities at home or lack transport, test scores may not improve even if high-quality teaching is available. A ToC that identifies regular participation as a key intermediate outcome would make this visible from the start, pointing to measures that support attendance, rather than simply expanding the tutoring offer.
- **Helping to clarify which indicators and data are needed** by naming the key steps and mechanisms. Instead of collecting many data points simply because they are available, teams can focus on a small set of indicators that correspond to the main links in the chain – from activities and outputs, through intermediate outcomes, to final outcomes. This makes monitoring more proportionate, reduces unnecessary reporting, and yields data that are easier to interpret and use.
- **Providing an important foundation for many evaluation approaches**. Impact evaluations, including experimental and quasi-experimental methods, are designed to test specific causal pathways. Without a shared ToC, it is difficult to design such evaluations in a credible and efficient way. Strengthening the use of ToCs can therefore support, where relevant, the possibility to generate stronger evidence on what works, for whom, and under what conditions.

In summary, using ToCs more systematically can help ensure that the substantial resources behind Programme "Education" are directed towards activities that have a real chance of improving outcomes, that emerging problems are detected in time, and that decision makers have a clearer basis for choosing between competing uses of scarce funds.

Practical limitations of using theories of change

ToCs are a useful organising tool, but they do not resolve all evaluation challenges on their own. By itself, a ToC does not demonstrate that the intervention is addressing the most relevant problem. For that reason,

it is most useful when combined with a sound needs assessment, stakeholder consultation and, where possible, existing evidence on the scale and distribution of needs.

Furthermore, a ToC necessarily simplifies reality. This is part of its value, as it makes complex interventions easier to discuss and manage. At the same time, there is a risk of presenting change as more linear, certain and mechanical than it will be in practice. In education, the same activity may work differently across institutions, learner groups or local contexts, and some effects may become visible only during implementation. A ToC should therefore be treated as a structured starting point for reflection, rather than as a precise prediction of what will happen. It is also important to strike the right level of detail: if a ToC is too broad, it may remain vague and of limited practical use; if it is too detailed, it can become difficult to apply and may place unnecessary demands on those expected to develop and use it.

There are also practical considerations. Developing and using a good ToC requires time, facilitation and a degree of evaluative capacity. If it is prepared only for an application or a report, and is not revisited thereafter, it can become a static document with limited value for implementation. To be useful, the causal logic needs to be explicit, shared and reviewed regularly, especially when monitoring points to bottlenecks, weak uptake or unintended effects. At beneficiary level, there is also a risk that incentives shape how ToCs are formulated. If applicants believe that only highly linear, optimistic or "risk-free" causal chains will be viewed favourably, this may discourage open discussion of uncertainty, implementation constraints or possible unintended effects.

Finally, a ToC does not answer questions of attribution: it does not show whether empirically observed changes were caused by the intervention. Rather, it helps clarify what the intervention is expected to achieve, through which mechanisms, and what evidence would be needed to assess whether this is happening. In this sense, a ToC is best understood as a living analytical framework for design, monitoring and learning, which can also support stronger evaluation.

2.2. Current status and challenges for theories of change in Programme "Education"

Overall, document review and interviews suggest that Programme "Education" already contains many of the elements required to develop a ToC – clear high-level goals, defined procedures and a comprehensive indicator framework. However, these elements have not yet been combined into an explicit causal chain. As a result, it is sometimes difficult to see how funded activities are expected to deliver progress towards programme objectives, and whether such progress materialises.

Programme "Education" provides a strong starting point for a more systematic use of theories of change

Programme "Education" documents and guidelines for applicants are very clear about what the programme is trying to achieve. They describe, in detail, challenges such as low participation in pre-school, high rates of early school leaving, uneven quality across schools, teacher shortages, limited inclusion of vulnerable groups, and weak labour market outcomes for some VET and higher education graduates. The corresponding goals (e.g. "improve labour market relevance of VET and higher education") are well articulated at a qualitative level.

At the procedure level, guidelines for applicants specify specific objectives and criteria documents list the activities that can be funded. Recently, the *Modernisation of Higher Education Institutions 2.0* procedure introduced a simple logical matrix in which applicants list key challenges, proposed solutions, activities and target groups (EAPE, 2025[4]). This is a significant step towards creating a more structured way of presenting ideas for projects that can contribute to achieving specified objectives.

The programme is also supported by a formal monitoring framework that is aligned with the European Commission's Common Provisions Regulation and European Social Fund Plus (ESF+) regulations. Output and result indicators, targets and reporting requirements are defined in line with European Union (EU) rules, and the Executive Agency has developed detailed application and reporting templates.

This means that Programme "Education" already documents goals, target groups and activities in a relatively systematic way. These are natural inputs for ToCs at programme, procedure and project level.

The causal pathways linking goals, activities and indicators could be clearer

The methodology and selection criteria developed under each Programme "Education" operation identify which types of activities are eligible for support and provide some foundation for understanding how these activities could contribute to change (EAPE, 2025[5]). However, it is not always clear how and why these actions are expected to produce change, through which steps, and under what assumptions. In the absence of such an explicit hypothesised causal pathway, it becomes difficult to judge whether an activity is the most promising way to achieve a given goal.

This also limits the scope for developing a shared ToC at procedure or project level. Specific objectives, activities and indicators sit side by side in Programme "Education" documentation, but the *"if we do X, we expect Y to change, leading to Z for the whole target group"* causal chain remains largely implicit. As a result, it is harder for programme actors to build a common understanding of how particular activities are expected to influence outcomes of interest, through which mechanisms, and under what conditions.

Potential constraints and unintended effects could be brought more clearly into the intervention logic

A second challenge is that constraints and potential unintended effects are not always made explicit in the design logic. For example, the guidelines for applicants for *Modernisation of Higher Education Institutions 2.0* note that 85% of Master's graduates end up working in graduate positions (associated with higher incomes), while only 51% of Bachelor's graduates work in similar positions (EAPE, 2025[4]). One possible response would be to encourage more students to get Master's degrees. However, this would be short-sighted: if the number of attractive graduate jobs in the labour market does not change, more Master's graduates will simply mean more applicants competing for the same positions. Some will end up in jobs that do not require advanced qualifications, meaning they will be overqualified, and others may be crowded out. In this case, the activity may change who holds which degree without improving overall employment outcomes. A ToC would require this labour market risk to be explicitly identified and discussed.

Similarly, encouraging student mobility is an important objective of the procedure, but in fields with very heavy curricula, such as medicine, students may not be able to take up mobility opportunities without adjustments to workload and course sequencing. Technical reports submitted by medical universities for the *Modernisation of Higher Education Institutions* procedure under the previous programming period document very limited uptake of mobility programmes, noting "due to exceptional workload, medical students cannot afford a three-months stay abroad, therefore no candidates were found to carry out this activity" (EAPE, 2023[6]). Without a ToC that takes such key constraints into account, mobility schemes may mainly benefit a small group of specific students and have limited effect on the wider target group.

The indicator framework could be more informative about progress towards goals and the mechanisms behind it

A third challenge is that the current indicator framework does not always provide information on whether expected mechanisms are operating and whether progress is being made towards intermediate or final goals. Most indicators used in Programme "Education" track direct outputs, such as, number of

participants, trainings delivered or institutions supported. These are important for accountability, but they do not show whether behaviours, practices or outcomes are changing.

Programme “Education” already uses several “result indicators”. However, there is scope to strengthen their value as indicators of programme impact. Currently, these indicators are often:

- **Participant based** – they count how many participants experienced a given improvement after the supported intervention. This improvement is difficult to interpret without information on the share of similar non-participants who improved over the same period.
- **Complex in formulation** – they may combine participation status, change and timing in a single measure, for example.
- **Not anchored in a clear denominator that reflects the whole reference group** – this could include, for example, all Grade 8 students in a school, or all VET graduates in a region.

Such indicators provide only limited information on whether or to what extent the programme itself contributed to an observed change. For instance, knowing that a certain number of vulnerable participants improved their results one year after an operation (as monitored in *Success for You* (EAPE, 2023[7]), for example) does not show whether the project itself caused the observed improvement. The same pattern could arise for at least two reasons:

- Outcomes may have improved for the wider group over the same period due to broader changes in schools, teaching conditions or other contextual developments.
- Those who choose to participate, or have the capacity to do so, may differ from non-participants in ways that matter for later outcomes, for example in motivation, confidence, access to information or readiness to engage. They may therefore have been more likely to improve even in the absence of the project.

In addition, such indicators do not show whether the project generated wider positive or negative effects on non-participants, for example through group dynamics, information flows or displacement. Nor do they show whether the mechanisms expected to drive change, such as more effective teaching practices, better attendance or stronger employer engagement, operated as intended.

Interviews with school leaders and non-governmental organisations suggest that there is a more basic bottleneck: difficulty in quantifying qualitative goals. When goals are phrased in broad terms – for example *“*increase labour market relevance of VET and higher education*”* – the best way to measure these may not be obvious to those defining indicator frameworks. This challenge is reinforced by the ESF+ indicator guidance (European Commission, 2021[8]), which places strong emphasis on participant-based indicators and offers relatively little guidance on how to operationalise concepts such as relevance or quality at system level. In this context, it is understandable that attention shifts back to counting participants and activities, which feel more concrete. A ToC can help by encouraging a discussion of what goals like “labour market relevance” mean in practice and by identifying feasible quantitative indicators for these elements.

Taken together, these three challenges have direct implications for the potential use of ToCs in Programme “Education”. When hypothesised pathways from activities to outcomes remain implicit, when constraints and unintended effects are not brought clearly into the intervention logic, and when indicators do not track mechanisms and outcomes in a way that supports interpretation, it becomes harder to use evidence to refine implementation and keep reform objectives in view. One interviewee talking in the context of VET reforms, summarised this risk succinctly:

> *“You want to implement a reform through a project, and you end up implementing the project without the reform.” – Employer association representative*

This comment reflects what several stakeholders and the evaluation team observed more broadly. When the causal pathway from reform goals, through mechanisms, to measurable outcomes is not made explicit,

day-to-day efforts naturally gravitate towards delivering the project activities and completing the required reports. Compliance with regulations and contractual agreements is essential for accountability and mandated by the European Commission. Whilst this must be upheld, when monitoring and evaluation concentrate exclusively on compliance, they risk overlooking whether activities are contributing to the intended outcomes and reform objectives.

2.3. Recommendations to strengthen the use of theories of change in Programme "Education" 2021-2027

The current challenges for ToCs in Programme "Education" do not reflect a lack of commitment to improvement. Rather, they suggest that there is room to introduce ToCs more systematically as a practical tool to connect existing goals, activities and indicators through explicit expected causal pathways, and to support a gradual shift from implementing projects towards tracking progress on reforms. The next section sets out practical recommendations and concrete options to make this shift happen.

Recommendation 1. Introduce simple theories of change at all levels

A first step is for the Executive Agency to include explicit ToCs at all key levels of the programme (i.e. procedure level, project level, activity or beneficiary level). The aim is not to produce long or technical documents, but to ensure that each level has a simple shared map of how its activities are expected to produce change.

At each level, it would be sufficient to develop a one-page ToC that sets out:

- **Inputs and activities** under direct control (e.g. teacher coaching visits, new learning materials, student career services).
- **Outputs** delivered (e.g. teachers trained, coaching visits completed, schools equipped, guidance sessions held).
- **Intermediate outcomes** that show whether the mechanism is working (e.g. observed changes in teaching practice, increased attendance among target groups, utilisation of resources provided, greater employer involvement in VET or higher education).
- **Final outcomes** that matter for learners and institutions (e.g. retention, completion, learning achievement, inclusion of vulnerable groups, employment and job quality for graduates).
- **Assumptions** that need to hold for the links from outputs to outcomes to work (e.g. stable staffing, realistic timetables, data access, buy in from school leaders and teachers, willingness of employers to cooperate).

These one-page ToCs can be annexed to programme documents, beneficiary guidelines and monitoring processes. Over time, they can also be used as a reference point in monitoring and review meetings.

Recommendation 2. Develop and test theories of change collaboratively

ToCs are most useful when they reflect the experience of those who design, implement and are affected by the intervention. It is therefore helpful if the Executive Agency can facilitate a process in which the ToC is developed in a collaborative way, involving:

- Procedure and project designers (e.g. Executive Agency, Ministry of Education and Science);
- Beneficiaries and implementers (e.g. schools, Centres of Vocational Excellence, higher education institutions, regional education departments, municipal administrations);
- Representatives of interest groups (e.g. students, parents, employers, professional associations).

Where it is not feasible to involve a broad group from the outset, ToCs could at least be stress-tested with key stakeholders. For example, a draft ToC for a new VET intervention could be discussed with employer representatives and VET teachers to check whether the assumed mechanisms and constraints are realistic. These discussions do not need to be formal or resource intensive. They can build on already existing opportunities for potential beneficiaries and stakeholders to provide feedback and be adapted to short preparation times for procedures. What is most important is that time is expressly allocated for this exchange so that the ToC reflects how the policy is likely to work in practice, not just on paper.

Recommendation 3. Link theory of change elements to indicators and plan data in advance

ToCs become more powerful when their elements are **linked to a small set of indicators**. A useful rule of thumb is that each box in the ToC should have at least one clear indicator, with:

- A defined data source (e.g. school administrative data, national assessments, graduate tracking data, surveys, employer feedback);
- A measurement window (e.g. end of semester, end of school year, 6 or 12 months after graduation).

When choosing indicators, three characteristics appear particularly important in the context of Programme "Education":

1. **Identify intermediate outcome indicators that can show progress in the short and medium term**. Examples include:
    - Measures of teaching practice (e.g. classroom observations, self-assessment, peer review).
    - Utilisation of resources (e.g. frequency of use of digital tools, uptake of support services, participation in work-based learning).
    - Simple student or module evaluation surveys in higher education (e.g. perceived relevance, clarity, workload, support).

    These intermediate indicators provide early signals about whether the mechanisms in the ToC are working long before final examination or employment data are available.

2. **Include a focus on cohorts or whole reference groups wherever possible.** It is helpful to define indicators for the full population of interest rather than only for participants. A simple rule of thumb is that the reference group should already exist in the year before the activity starts so that before–after comparisons are possible. For example:
    - Instead of only counting the number of vulnerable participants whose academic outcomes improved one year after participation, an indicator could track the percentage of vulnerable students who reach a defined proficiency level this year compared to last year. Reference groups can be refined as needed (e.g. Grade 4 students for whom Bulgarian is not their first language in a particular school) but should not condition on participation in specific activities (see Box 2.1).
    - Instead of focusing only on participants in a support scheme, an indicator could track the early school leaving rate of all students in the target schools, again disaggregated by reference group.
3. **Frame more indicators as percentages or shares, rather than as absolute counts only**. For example, use the share of students completing Grade 8 or the share of teachers using a new method rather than the number of completers or trained teachers. Absolute numbers are highly sensitive to the size of the school, region or programme and therefore difficult to compare across institutions or over time. Programme "Education" already has some indicators that are expressed

as percentages and this could be further developed as percentages and rates allow more meaningful comparisons with baselines, with other schools, and with potential reference or control groups.

The strong ESF+ focus on participant-based indicators will continue to be important for assessing compliance. However, experience from other EU Member States, such as Ireland, shows that Managing Authorities often complement ESF+ indicators with their own key performance indicators that are better aligned with their ToC and easier to interpret. Bulgaria has a rich existing framework of specific objectives, which could be aligned with strengthened ToCs. The availability, accessibility and quality of system-level data is a key consideration when identifying such indicators (see Chapter 4).

Box 2.1. Using indicators to make inferences about the causal impact of a programme

Given the effort that goes into collecting monitoring data, it is understandable and desirable to ask whether these indicators can also be used to say something about programme impact. In practice, however, such inference is complicated by the absence of a clear counterfactual: it is not observed what would have happened for the same learners, teachers or institutions in the absence of the programme.

One way to address this problem is through counterfactual impact evaluation, using experimental or quasi-experimental methods. However, these approaches are not feasible or proportionate in all contexts. Where they are not feasible or practical, carefully designed indicators can still provide a useful first proxy for impact, even if they do not, on their own, deliver a clean causal estimate.

In this context, changes in cohort-based indicators are often more informative than participant-only indicators. For example, a change in the share of all vulnerable students in a school who reach a defined proficiency level captures both how many students were reached and how outcomes changed in the target population as a whole. Such indicators do not prove that the programme caused the observed change since outcomes may also be influenced by other developments affecting the group over time, such as other reforms, demographic change, or broader contextual developments. These wider influences may complicate the interpretation, but they are often at least visible and can be discussed explicitly. By contrast, the common alternative of comparing participants with non-participants who appear similar on observed characteristics can be misleading, because those who choose to or have the capacity to participate may differ in less visible but important ways, such as motivation, confidence, access to information, or capacity to engage. These differences are often harder to observe and adjust for, even though they may strongly influence the outcomes of interest.

For this reason, where the aim is to obtain a first proxy for impact from routine monitoring, cohort-based indicators are often the more informative default. Stronger causal claims should rely on additional evaluation designs where feasible, complemented by qualitative evaluation methods. Participant-level indicators remain important for operational, financial and compliance monitoring.

Recommendation 4. Quantify qualitative goals through simple proxies

Many of the most important objectives of Programme “Education” are qualitative in nature, for example, *“*increasing labour market relevance of VET and higher education*”* or *“*improving the quality of teaching*”*. ToCs can help the Executive Agency and the Ministry of Education and Science to translate such qualitative goals into measurable elements through plausible proxies. These proxies will never capture the full richness of the underlying concept, but they can provide a reasonable indication of whether the system is moving in the right direction, especially when complemented with more in-depth evidence that explores how changes are experienced in context and helps interpret results. A ToC provides a structure within

which such proxies can be identified and linked to upstream mechanisms, such as curriculum changes, partnerships with employers and student guidance. For example:

- For **labour market relevance of VET**, proxies could include:
  - share of VET graduates employed in occupations related to their field of study;
  - employer satisfaction with the skills of recent VET graduates in key sectors;
  - extent of employer participation in curriculum design or work-based learning.
- For **labour market relevance of higher education**, proxies could include:
  - employment rate of graduates in jobs requiring a tertiary education qualification;
  - share of higher education graduates employed in occupations related to their field of study;
  - employer ratings of the preparedness of graduates on dimensions such as problem solving, communication or digital skills.

Quantifying qualitative goals inevitably means managing several constraints, including restrictions related to the protection of personal data, as well as current difficulties in the systematic collection and use of certain types of data in Bulgaria. Chapter 4 considers ways that the use of system-level data could be strengthened in the Bulgarian education system more generally. It could also draw on existing national system indicators and internationally comparable OECD Indicators of Education Systems Programme, adapting definitions to Bulgarian data availability.

Recommendation 5. Use theories of change to support learning and incentives, focussed on creating impact

Finally, ToCs can help the Executive Agency to support a gradual shift towards more incentives and learning, encouraging those who manage and implement projects to consider the impact of their projects more fully, without adding heavy administrative burdens:

- They can inform the design of targets and low-stakes incentives that place at least some weight on results outcomes. This relates to the broader question of outcome-linked financing, which is discussed further in Chapter 3.
- They can be used to enrich the existing dashboards with a small number of intermediate and final indicators along the causal chain, and to link these to simple feedback routines – for example, predefined thresholds trigger concrete follow-up actions or support - so that emerging problems are detected early and examples of good practice can be identified and disseminated.

In this way, ToCs become a practical bridge between the high-level ambitions of Programme “Education”, the concrete activities funded under its procedures, and the monitoring and evaluation tools that can help ensure that limited resources are used where they make the greatest difference.

References

EAPE (2025), *Application conditions under the procedure for direct award of grant aid BG05SFPR001-3.006 “Modernisation of Higher Education 2.0”*, https://eumis2020.government.bg/bg/s/Procedure/Info/5cdb7749-ad19-4917-b58c-f9003525ed68 (accessed on 22 December 2025). [4]

EAPE (2025), *Criteria and Methodology for Evaluation of Project proposal under the Direct grant procedure "Modernisation of Higher Education Institutions 2.0" (Annex X)*, https://files2020.government.bg/f381cc32-3b29-4260-8ae0-0a54fd5729d1?access_token=eyJhbGciOiJodHRwOi8vd3d3LnczLm9yZy8yMDAxLzA0L3htbGRzaWctbW9yZSNobWFjLXNoYTUxMiIsInR5cCI6IkpXVCJ9.eyJpc1VwbG9hZGVyIjoiRmFsc2UiLCJ0eXBlIjoiU2luZ2xlS2V5IiwiYmxvYktleSI6ImYzODFj (accessed on 22 May 2026). [5]

EAPE (2023), *Project implementation monitoring questionnaire (Annex XXI)*. [7]

EAPE (2023), *Technical report: Contract No. BG05M2OP001.2.016-0007-C01, Technical report No 3*. [6]

European Commission (2021), *ESF+ Common indicators index: Working document*, https://sfc.ec.europa.eu/system/files/documents/documents/toolbox-october-2021_0.pdf (accessed on 23 December 2025). [8]

Mayne, J. (2017), "Theory of Change Analysis: Building Robust Theories of Change", *Canadian Journal of Program Evaluation*, Vol. 32/2, pp. 155-173, https://doi.org/10.3138/cjpe.31122. [3]

Research, U. (ed.) (2014), "Theory of Change", *Methodological Briefs: Impact Evaluation No. 2*, https://www.entwicklung.at/fileadmin/user_upload/Dokumente/Evaluierung/Theory_of_Change/UNICEF_Theory_of_change.pdf (accessed on 13 May 2026). [2]

Weiss, C. (1995), *Nothing as practical as good theory: Exploring theory-based evaluation for comprehensive community initiatives for children and families*, Aspen Institute, Washington, DC., https://docs.opendeved.net/lib/2URBNM2X (accessed on 13 May 2026). [1]

3 Supporting evaluation thinking and planning among beneficiaries

The previous chapter focused on how Programme "Education" 2021-2027 could use theories of change (ToCs) more systematically at all levels. This chapter focuses specifically on the beneficiaries of Programme "Education" – such as the Ministry of Education and Science, higher education institutions (HEIs), municipalities, schools and Centres of Vocational Excellence – and asks how they could be further supported to use evaluation thinking in their work and to view reporting, monitoring and evaluation as a continuous learning process embedded from the design stage. First, it looks briefly at the current context of reporting, monitoring and evaluation in Programme "Education". Then it presents a set of recommendations and possible measures to support beneficiaries.

3.1. Current status and challenges for evaluation thinking among beneficiaries

Across the interviews with beneficiaries and with officials in research, analysis and monitoring functions, there was a broad consensus that monitoring and reporting currently focus mainly on compliance with regulatory, procedural and contractual requirements. This occurs at every stage of implementation:

- **Application stage:** Applicants under both direct and competitive procedures prepare extensive application packages that focus heavily on formal eligibility criteria and on documenting that all required elements are in place. For competitive calls, mathematical scoring models are used to rank proposals. These models tend to place substantial weight on output targets, in particular the planned number of participants. This can encourage a strong focus on delivery plans and participant volumes while giving less structured attention to whether the proposed activities are likely to generate the intended outcomes and how this would later be recognised in practice.
- **Assessment and contracting:** For direct awards, assessment grids focus almost exclusively on formal eligibility; for competitive calls, scoring mostly compares the volume of planned activity and the extent to which eligibility conditions are met. The quality of the underlying intervention logic or ToC plays a smaller role.
- **Reporting:** Once a grant is awarded, beneficiaries face a heavy reporting burden. At least every six months, they prepare technical and financial reports often accompanied by numerous annexes. These reports are essential for accountability and many of the requirements stem from European Social Fund Plus (ESF+) regulations. In practice, the reporting framework is largely procedural in focus, concentrating mainly on whether the planned number of participants was reached, scheduled activities were delivered, and the required documentation is in place to justify expenditure and demonstrate compliance with rules.

  One beneficiary summarised this experience by noting that:

> *"The reporting requirements are too far up front and too detailed – we need to report in February who will be the speakers at an event in November".*

Whilst in practice technical and financial reports cover activities that have already taken place and not future predictions, this quotation shows that many beneficiaries experience monitoring and evaluation primarily as a matter of administrative control rather than a means of understanding and improving the impact of their work.

The Executive Agency Programme "Education" (hereafter, Executive Agency) has already taken meaningful steps to move beyond a sole focus on compliance. In the current programming period (2021-2027), it has introduced monitoring questionnaires for most procedures, asking beneficiaries to report about progress towards policy goals. This represents an important signal of commitment to quality and a willingness to innovate in monitoring practice. While it is still too early to assess their full effects, interviews and document review suggest several aspects where their current functioning appears more limited than their potential:

- At present, the statistics elicited in the questionnaires inherit many of the issues described in Chapter 2 (e.g. they are often participant-based and difficult to interpret in terms of impact).
- Responses are not systematically tied to beneficiary incentives, and do not trigger formal feedback or learning, support or correction mechanisms. As a result, beneficiaries have little direct reason to invest effort in achieving policy objectives, or even in the quality of the information provided.
- For some beneficiaries, the questionnaires are experienced as an additional reporting requirement rather than as a tool for learning. This suggests that the intended purpose – supporting learning and signalling strategic priorities – may not yet be fully apparent to all users.

Taken together, this pattern makes it easy for beneficiaries to lose sight of the quality and impact of the activities they deliver. The main measure of success becomes *"*have we implemented what we promised*"* rather than *"*has anything improved for the learners and groups we serve*"*.

Why support evaluation thinking among beneficiaries?

Given this context, this report proposes to strengthen opportunities for beneficiaries to engage with evaluation thinking in a more deliberate way. The starting point is simple: beneficiaries are closest to the design and delivery of activities, and they are often best placed to see what works, for whom and under what conditions. If they are only asked to demonstrate procedural compliance, their practical knowledge is not fully used.

Encouraging beneficiaries to adopt evaluation thinking from the start of a project has several potential benefits:

- It prompts teams to think through causal pathways at the design stage – who is the target group, what exactly will change for them, through which mechanisms, and how does this connect to the wider goals of Programme "Education". This links directly to the use of ToCs as described in Chapter 2.
- It helps beneficiaries to focus on quality and impact rather than only on volume. For example, a school or HEI is encouraged not only to recruit participants and monitor participant outcomes including satisfaction, but also to ask whether practices are changing, whether support reaches the right groups, and whether outcomes for the whole reference group are improving.
- It supports better services for learners. When teams regularly reflect on what is working and why, they are more likely to identify bottlenecks early (e.g., low uptake by certain groups, weak implementation in some schools, or unintended effects) and to adjust their approach.
- It helps to create shared understandings between stakeholders at different levels. For instance, if the Executive Agency, the Ministry of Education and Science, a Centre of Vocational Excellence and partnered vocational education and training (VET) schools all work with simple and

interconnected ToCs, it becomes easier to ensure that expectations are coherent and that data collected at different levels can be interpreted in a consistent way.

At present, because causal and impact questions are rarely addressed in internal grant documents, evaluation is often viewed as a purely externalised activity. Internal actors focus on delivering activities and submitting reports while questions about mechanisms, intermediate outcomes and impact are largely left to external evaluations that are often carried out with some delay. This separation makes it difficult to build simple evaluation thinking into everyday programme management (e.g. checking key assumptions, monitoring a small set of intermediate indicators, or documenting variation in delivery). Strengthening these internal habits is not a substitute for independent evaluation, which remains essential, but a way to ensure that external evaluations are complemented by ongoing learning within programmes.

The remainder of this chapter sets out practical support measures that the Executive Agency could integrate into its procedures to help beneficiaries use evaluation thinking as part of their regular work.

3.2. Recommendations for support measures for beneficiaries

Recommendation 1. Require a theory of change at the application stage

To bring evaluation thinking closer to beneficiaries, the Executive Agency could require that each application includes a short ToC from the outset. This is consistent with international guidance that recommends making the results chain explicit at design stage so that implementation and monitoring can focus on outcomes and not only on activities (OECD, 2021[1]; Vogel and Isabel, 2012[2]; Khattri and Roberts, 2012[3]).

- **Ask for a one-page ToC.** Every applicant (direct and competitive procedures) would submit a one page ToC that sets out, in simple form: key activities and inputs, immediate outputs, a small number of intermediate outcomes, final outcomes for the target group and two or three main assumptions. The format should be a basic diagram to illustrate key links between elements.
- **Provide a standard template and example.** The Executive Agency could supply a short template with these elements pre-defined, plus one worked example from an existing Programme "Education" procedure (e.g. *Success for You* or a project under *Supporting Centres of Vocational Excellence*). A link to a simple public ToC guide or tool (such as a free online theory of change builder – e.g. TASO (2025[4])) would be sufficient for those who want more detail.
- **Ask beneficiaries to propose indicators.** The template could build on existing practice by asking applicants to propose at least one indicator (with data source and timing) for each key element of the ToC, especially for intermediate and final outcomes. This prepares later work on targets, monitoring and data access.
- **Use ToCs in competitive assessment.** For competitive calls, the quality and plausibility of the ToC could become an explicit criterion in the scoring grid. Reviewers would look at whether the proposal sets out a credible chain from activities to intermediate and final outcomes and recognises key risks and assumptions. Proposals with a more thoughtful ToC would receive higher scores, signalling that understanding mechanisms and pathways to impact is valued alongside participant numbers and budgets. This approach is consistent with ESF+ guidance that stresses the importance of a clear intervention logic as an essential prerequisite for assessing programme proposals (European Commission, 2021[5]).
- **Offer ToC and indicator workshops to support indicator use.** Before application deadlines, the Executive Agency could run short online ToC and indicator workshops where applicants can ask questions and get feedback on their draft ToC and on how it links to the procedure's indicator framework. This could be an additional item in the information sessions that are already held with

beneficiaries. At contracting stage, Executive Agency staff and beneficiaries can then agree which indicators from the existing programme or procedure-level framework are most relevant for their work package and specify clear numerical targets, target bands and reporting arrangements for those selected indicators. Keeping the selected indicator subset small but meaningful helps to ensure that indicators are aligned with the causal chain and that both the Executive Agency and the beneficiary have a shared understanding of what success would look like. It also prepares the ground for the later support measures on targets, internal monitoring routines and the creation of incentives discussed in Recommendations 2 and 6.

Recommendation 2. Use targets as planning tools

In the current programming period, the Executive Agency has already made substantial progress on indicator design and target setting. With support from the World Bank, Bulgaria has introduced a structured methodology that links micro-indicators at operation level to programme-level indicators, uses historical data and unit costs to derive realistic targets, defines milestones for 2024 and end-period targets, and documents assumptions and risks in detailed indicator factsheets. This is fully aligned with ESF+ expectations on baselines, milestones and targets (EAPE, 2022[6]). This is a solid foundation that is consistent with ESF+ guidance on evidence-based baselines, milestones and targets.

A valuable next step could be to use these targets more explicitly as practical planning tools in the dialogue between the Executive Agency and beneficiaries, especially around the six-monthly technical and financial reports that underpin intermediate and final payments. The proposals on target ranges and follow-up below apply mainly to programme-specific indicators and key project-level indicators. ESF+ common indicators will continue to be used for compliance and reporting to the European Union (EU) but need not be the only focus of planning discussions.

- **Clarify the role of targets in Executive Agency-beneficiary dialogue.** Targets for key indicators – whether they are ESF+ common indicators or programme-specific indicators – can be presented to beneficiaries not only as formal commitments but as shared expectations about what will happen over time. In practice, this means that six-monthly reporting would routinely compare the current value to the agreed path and use this comparison to structure the discussion: are we broadly where we expected to be, a little behind, or seriously off track, and why?
- **Use target ranges to trigger proportionate follow up.** Without changing the single milestone and final target reported to the European Commission, the Executive Agency could internally distinguish three levels for key indicators in its discussion with beneficiaries:
  - a **target floor** – a minimum acceptable level;
  - the **target** – the expected level, in line with the indicator factsheet;
  - an **ambitious target** – a level that would indicate particularly strong performance.

  When six-monthly reports are reviewed:

  - **At or above the target**: the project is considered on track. If results are close to the ambitious target for several periods, the Executive Agency may invite the beneficiary to share its approach with others.
  - **Between the target floor and the target**: the Executive Agency and beneficiary agree one or two light-touch corrective actions, such as small adjustments in outreach, scheduling or implementation support.
  - **Below the target floor**: the Executive Agency and beneficiary jointly develop a more structured support plan, such as more frequent follow up, coaching visits, revisiting key design choices, or, in some cases, revising the target if original assumptions clearly no longer hold.

In this way, target ranges help to translate indicator values into concrete feedback loops and support plans, rather than serving only as pass-fail thresholds. At the aggregate level, this information can also be used to improve selection. For example, it can help identify the characteristics of projects and partnerships that have been more successful or to ensure that targets are not unrealistic at the proposal stage.

- **Ask beneficiaries to document how they derived their project-level targets.** When beneficiaries propose their own targets for project-level indicators, which later feed into programme-level ESF+ common and programme-specific indicators, application and contracting templates could invite a short explanation of the underlying reasoning. For example, beneficiaries could refer to:
  - local baselines (e.g. recent values for the same cohort or school);
  - expected participation and completion under realistic assumptions;
  - capacity constraints (e.g. staff, rooms, timetable, employer availability).

  This does not replace the existing programme-level methodology, which already incorporates baselines, capacity and trajectories. Instead, it extends the same logic downwards and makes later discussions about under- or over-performance more constructive: the focus moves to which assumption did not hold, and what can be adjusted, rather than simply whether a numeric target was met.
- **Encourage more outcome- and cohort-based programme-specific indicators**. Finally, where feasible, the Executive Agency could gradually expand the set of programme-specific indicators to include more outcome- and cohort-based measures that align closely with the ToCs described in Chapter 2. Examples include:
  - The share of students from specific vulnerable groups in a given grade reaching a defined proficiency level;
  - Completion and early-leaving rates for all students in schools participating in an intervention, compared with their own baselines;
  - Employment rates of vocational or higher education graduates in related occupations after 6-12 months, or earning above a certain threshold, by field and region.

  Targets and target ranges for such indicators can then be used in Executive Agency-beneficiary dialogue to complement ESF+ common indicators. This would allow Programme “Education” to retain full compliance with ESF+ requirements while using a richer set of indicators to plan, monitor and adjust interventions in line with its own policy goals.

Recommendation 3. Further simplify compliance aspects of reporting and embed more learning

Bulgaria has already taken important steps to simplify financial control. Interviews and documents indicate that Simplified Cost Options (SCOs) are widely used and that the EU Management Information System (EUMIS) is the common platform for uploading and tracking project documentation. This is fully in line with ESF+ expectations on proportional monitoring and use of simplified costs.

Building on this base, two complementary directions could be considered: a further review of compliance reporting to identify where simplification is still possible and the addition of a concise learning element to regular reports.

- **Keep the compliance pack but simplify, where possible:**
  - **Maintain core compliance reporting** but continue to review whether some elements can be shortened, merged or moved to sample-based checks. Given that SCOs are already in place,

a focused review could examine whether all current annexes and data requirements remain necessary for operations using SCOs or whether some checks can be reduced or shifted towards ex-post audits on a smaller sample of projects.

- **Use EUMIS to streamline rather than multiply requirements.** Where feasible and not already in place, reporting templates and uploads could be standardised within EUMIS so that beneficiaries do not need to submit the same information multiple times in different formats.
- **Review which data requirements are legally mandatory, and which are guidance.** The ESF+ Common Indicators Toolbox explicitly notes that for some common output indicators, "data collection is only necessary when applicable and, in relevant cases, informed estimates can be used instead" (European Commission, 2021[5]). This suggests that Bulgaria may have some discretion in how far it goes beyond the minimum requirements. A structured review of indicator factsheets and national templates could identify where national practice is stricter than EU rules require, and where proportional approaches (e.g. sampling, informed estimates, or reduced breakdowns) would be acceptable.
- **Consider gradual expansion of Financing Not Linked to Costs (FNLC)** where appropriate. For selected, well-defined results (e.g. dropout rates of vulnerable groups in a project like *Success for You*), FNLC can reduce the volume of cost documentation per participant and shift checks to whether agreed conditions are met. This may free up some administrative capacity on both sides, which can be redirected to learning-oriented activities.

Given the complexity of ESF+ regulations, any further simplification will need careful legal analysis. The key recommendation is therefore to review existing compliance reporting systematically, with ESF+ proportionality principles in mind, rather than to prescribe specific reductions.

- **Add a concise learning pack to regular reporting, linked to the theory of change**. In parallel, the Executive Agency could introduce a short learning component to the six-monthly reports. The aim is not to increase the total reporting burden, but to rebalance it towards reflection on quality, mechanisms and impact, using the project's ToC as a reference point. A possible structure would be:
  - **Dashboard section** – using existing EUMIS data, highlight a small number of intermediate indicators that correspond to key links in the ToC (e.g. changes in teaching practice, use of new resources, school attendance of target groups, early dropout patterns). This makes it possible to see whether the expected mechanisms are beginning to operate in practice.
  - **"What we learned" section** (approximately one page) – a short narrative where the beneficiary reflects explicitly against the ToC:
    - which parts of the causal chain appear to be working as expected;
    - where results are weaker than anticipated;
    - which assumptions may have proved unrealistic (e.g. on stakeholder buy-in, capacity, timetables or data availability);
    - whether any unintended effects or bottlenecks have emerged.
  - **"Next steps" section** (approximately one page) – a concrete list of two or three actions for the next period, again linked to the ToC (e.g. strengthening a weak link in the chain, addressing a newly visible bottleneck, adjusting an assumption) and with named owners.

The intention is to keep the overall effort broadly constant: wherever a new learning element is added, the Executive Agency would look for at least one compliance-related element that can be simplified, consolidated or moved to sample-based checks. Over time, this would help beneficiaries and the Executive Agency to use the ToC not only as a design tool, but also as a structured backbone for ongoing monitoring, feedback and course correction.

Recommendation 4. Encourage internal monitoring routines with feedback loops

At present, procedures under Programme “Education” set requirements for external reporting to the Executive Agency but say little about how beneficiaries might organise their own internal monitoring. This respects institutional autonomy but also means many beneficiaries receive limited guidance on how to use data for day-to-day implementation improvement. International guidance on results-based monitoring and evaluation stresses that routine internal monitoring of a small number of meaningful indicators, linked to the intervention logic, is essential for management and learning, not only for accountability (Zall Kusek and Rist, 2004[7]). As the focus on impact and, potentially, results-based financing increases (see Recommendation 6), internal monitoring routines can become an important operational lever for quality.

The Executive Agency could therefore encourage beneficiaries to establish simple, ToC-based monitoring habits that remain internal to the institution and do not create additional reporting obligations.

- **Monitor key intermediate outcomes at higher frequency**. Using the ToC as a guide, beneficiaries can identify a small number of intermediate outcomes that are easily observed at monthly or semesterly intervals. This might include changes in teaching practices, use of new learning resources, attendance of target groups, early signs of dropout, or student feedback on new modules. World Bank guidance on results frameworks underlines the value of such intermediate results as early-warning signals that inform management decisions before final outcomes are visible (Khattri and Roberts, 2012[3]). Tracking these indicators internally helps teams see whether the mechanisms in their ToC are starting to operate as intended.
- **Apply internal target bands with clear triggers.** For these intermediate indicators, institutions can set their own internal target floor, target, and ambitious target, and link them to simple traffic-light rules. For example, remaining between the target floor and target could trigger small local adjustments like changes in scheduling or communication. Consistently falling below the target floor might prompt a more substantial review within the institution. These bands would not need to be reported to the Executive Agency; they would be an internal management tool for the beneficiary.
- **Use a simple problem-solving method when challenges arise.** When internal data show that an indicator is repeatedly off track, beneficiaries can be encouraged to use a light problem-driven approach, such as Problem-Driven Iterative Adaptation (PDIA). PDIA involves defining the problem clearly, breaking it down into underlying causes, identifying feasible entry points for action, testing one or two small changes, checking whether the data improve, and then iterating (Andrews, Lant and Michael, 2012[8]). The aim is not to introduce a formal methodology, but to normalise the idea that monitoring data should trigger reflection and small adjustments rather than remaining purely descriptive or unused.
- **Review the training provided to staff from the Executive Agency who are contact points with beneficiaries to ensure they are able to act as learning officers and broker solutions.** It is important that contact with staff from the Executive Agency is not just focused on helping beneficiaries reach and document compliance, but that beneficiaries have someone they can contact to discuss substantive challenges, hear how similar problems have been solved elsewhere, and receive basic support on interpreting their own data. Recent OECD work on public policy evaluation highlights the value of “knowledge officers” who connect evidence producers and users and help foster a culture of learning across the administration (OECD, 2025[9]; 2022[10]). The role of contact points within the Executive Agency could be more explicitly connected to applied, evaluation-informed learning. This would involve providing protected time and targeted training to enable them to serve as a focal point for monitoring, evaluation and learning who remains up to date with developments in evaluation practice and evidence use.

These measures keep the responsibility for internal monitoring squarely with beneficiaries, while signalling that the Executive Agency values simple, ToC-informed feedback loops that help institutions learn from their own data and improve the quality and impact of their activities over time.

Recommendation 5. Facilitate peer learning and coaching

Strengthening evaluation thinking among beneficiaries should not rely only on written guidance. It is greatly helped when institutions can learn from each other's practical experience. Earlier OECD work evaluating the *Modernisation of Higher Education Institutions* procedure in Bulgaria already recommended that authorities organise peer learning among beneficiaries so that they can share common difficulties and solution approaches (OECD, 2025[11]). This report builds on that recommendation and suggests how similar peer learning mechanisms could support evaluation thinking more broadly across Programme "Education".

- **Facilitate communities of practice focused on implementation challenges.** The Executive Agency could convene regular, light-touch communities of practice for institutions working within the same procedure or project. Under direct grants where the beneficiary is the Ministry of Education and Science, this logic extends to groups of institutions (e.g. kindergartens, primary and secondary schools, VET schools, HEIs) working on similar activities. International evidence shows that structured networks and collaborative spaces make it easier for practitioners to discuss evidence, compare experiences and adapt approaches to local contexts (OECD, 2022[10]; OECD, 2023[12]). In practical terms, this could mean organising short quarterly workshops (around 90 minutes, online or regional) where project teams briefly present a challenge they are facing and how they have tried to address it and hear how others have approached similar issues. These sessions would focus on implementation and outcomes (i.e. motivating teachers to use new digital tools, or increasing participation of vulnerable students), rather than on administrative or financial compliance. To support open discussion, the Executive Agency (in the case of competitive procedures) or main beneficiary would retain responsibility for organising and framing these communities of practice, while the meetings themselves would be held without staff from the Agency in the room. Under some ongoing projects, structures may already be in place around which to organise these communities of practice. For example, under the *Bright Start* and *Success for You* projects, the OECD has previously identified that there are regular opportunities for project teams at national, district and institution level to meet and exchange (OECD, 2026[13]). In the future, dedicating some of these to workshops to implementation and outcomes could be helpful, as well as ensuring that workshops include opportunities for teams to exchange with each other rather than simply listen to presentations.
- **Provide a simple shared repository of practical resources.** To complement these exchanges, the Executive Agency could host and curate a concise online repository where beneficiaries upload short notes and tools that others can reuse. For example, this might include frequently asked questions, indicator dictionaries, short "how-to" notes on monitoring intermediate outcomes, or checklists for preparing ToCs. Similar repositories have been highlighted in OECD work on research use to support knowledge mobilisation and reduce duplication of effort (OECD, 2022[10]; 2023[12]). The aim would not be to create a complex new platform, but to offer a single, easy-to-access repository where practical resources are collected and updated.
- **Capture and share short "data stories".** Interviews with managing authorities in other countries suggest that peer learning works best when it is grounded in concrete experience and examples. One simple way to encourage this is to ask beneficiaries to submit a brief "data story" after each reporting cycle that could be shared in the community of practice. This could be as simple as one chart from their monitoring data plus two or three sentences explaining what changed, why it changed, and what they learned. Alternatively, it could be a short video recording sharing experiences and insights gained from data. For example, a school might show how attendance of

a target group evolved over a term, note a change in communication with parents, and explain how this shaped their next steps. These brief stories can help normalise the idea that monitoring data are not only for compliance but also for reflection on mechanisms and for small course corrections.

- **Use PDIA as a structured problem-solving approach.** When monitoring data or experience in the community of practice reveal persistent problems, beneficiaries could be encouraged and coached to use a light version of PDIA. The PDIA toolkit, developed by the Building State Capability programme at Harvard, offers a structured but flexible way of dealing with implementation challenges (Harvard Kennedy School, 2025[14]; Andrews, Lant and Michael, 2012[8]):
  - First, teams **construct the problem**, in plain language from the perspective of those affected. For example, "students are not using the new learning platform".
  - Second, they **deconstruct the problem** into causes and sub-causes. For example, access to devices, teacher confidence, time in the timetable or perceived relevance for examinations.
  - Third, they **identify feasible entry points** for action where there is overlap between authority (who can decide), acceptance (who will support the change) and ability (who has the skills and resources).
  - Finally, they **test small tweaks**, monitor intermediate outcomes, such as log-in rates or teacher use in class, and adapt based on what the data show over successive iterations.

The Executive Agency could introduce PDIA through guidance notes, examples drawn from procedures within Programme "Education", or optional training sessions, while communities of practice and individual institutions then apply PDIA in their own meetings, sometimes without staff from the Executive Agency present.

- **Clarify the role of the Executive Agency as the facilitator of the peer learning process.** Care should be taken to ensure that the Executive Agency is not perceived as judging beneficiaries or projects during peer learning, but rather as supporting reflection and improvement. Communities of practice, repositories and PDIA-style coaching are intended to create a safe space where beneficiaries can be honest about difficulties, share partial results, and jointly explore how to improve. This aligns with wider OECD work that stresses the value of "learning-oriented" infrastructures and intermediaries that help education systems move from compliance-based monitoring towards continuous, adaptive improvement (OECD, 2022[10]; 2023[12]).

Over time, such peer learning and coaching structures can help embed evaluation thinking in day-to-day practice: beneficiaries become more used to articulating their ToC, checking whether the key steps in the causal chain are happening and using simple data to refine their approaches, rather than seeing evaluation as something that happens only at the end of a project or only in external studies.

Recommendation 6. Calibrate indicators to align beneficiaries' incentives with policy goals

At present, almost all funding to beneficiaries under Programme "Education" is provided through cost-based reimbursement logic, mainly via SCOs such as flat rates, standard scales of unit costs and lump sums. This approach brings important benefits in terms of simplification and reduced administrative burden. At the same time, it means that payments depend on whether activities were implemented, and costs incurred, rather than on whether outcomes improved. As a result, beneficiaries receive very little financial encouragement to focus on the quality and impact of their activities. In practice, the programme therefore relies on their intrinsic motivation and professional values to ensure that what they do makes a difference.

One way to better align the incentives of beneficiaries with the policy objectives of Programme "Education" is to link a small share of funding to clearly defined results. The European Commission and the World Bank

now provide extensive guidance and practical examples on FNLC, also referred to as Results-Based Financing, including advice on how to design, sequence, and calibrate result-based payments (European Commission, 2021[15]; 2025[16]; World Bank, 2017[17]; 2021[18]). Box 3.1 provides an illustration from an ESF+ programme in Ireland. Taken together, the guidance and example demonstrate that FNLC has become a practical, widely used tool, including in education programmes (World Bank, 2017[17]; 2021[18]).

Box 3.1. Monitoring and evaluation of European Social Fund Plus projects in Ireland

Ireland offers a useful illustration of how an education and training system can organise evaluation work so that it supports both accountability and learning. In Ireland's European Social Fund Plus (ESF+) Employment Inclusion Skills and Training Programme, the Managing Authority's evaluation plan sets out priorities, timing, and a clear governance process for evaluations, including the use of evaluation steering groups and formal management responses to findings. This type of set-up can help ensure that evaluations remain closely connected to programme decisions (e.g. adjustments to delivery or targeting) and that findings are followed up systematically (Government of Ireland, 2023[19]).

Irish documentation demonstrates how evaluations in Ireland have combined theory-based approaches (to clarify mechanisms and implementation conditions) with some experimental impact evaluations, where this has been feasible. For example, Ireland has commissioned quantitative impact evaluations of ESF-funded training and education measures using quasi-experimental techniques alongside process analysis (Indecon, 2020[20]). This includes evaluations of further education and training programmes that rely on matching and weighted regression methods to estimate impacts on labour market outcomes.

A recurring enabling factor is the use of linked administrative data, which reduces reliance on bespoke beneficiary surveys. Several Irish evaluations link programme administrative records to longitudinal labour market datasets (such as the Jobseekers Longitudinal Dataset), allowing follow-up on employment and earnings outcomes over time (Indecon, 2020[20]). Recent improvements in national data infrastructure further expand the scope for counterfactual impact evaluation and longer-term follow-up, notably through the Central Statistics Office's Educational Longitudinal Database and the Department of Social Protection's Work and Welfare Longitudinal Dataset (Kelly et al., 2025[21]). Where legal gateways and governance arrangements allow, these same linkable infrastructures can also strengthen ESF+ evaluation by enabling participant trajectories and outcomes to be tracked beyond the end of funded support without creating new large-scale surveys (Kelly et al., 2025[21]).

Ireland's Momentum Programme provides a useful illustration of how results-based financing can be used to strengthen implementer incentives around delivery quality and meaningful participant outcomes.

An example of ESF+ results-based financing: The Momentum Programme (2012 – 2017)

Momentum funded free education and training projects for unemployed young people, as well as for long-term unemployed jobseekers, aiming to develop labour market-relevant skills and support progression into employment in growing sectors. Training was accredited to industry standards and combined with supported work placements. The programme was implemented in two iterations with Momentum 2 explicitly adjusted based on learning from Momentum 1 (SOLAS, 2014[22]; 2016[23]; OECD, 2020[24]).

A distinctive feature was the hybrid output-and-results-based payment model. Under Momentum 2, the structured payment model linked provider payments to both training delivery and participant outcomes, with a 70:30 split between skills delivery and work outcomes. Within the outcomes component, the

programme required measurable milestones including skills development, work placement, and a minimum period of at least two consecutive months of employment following placement.

This structure can be interpreted as a risk-sharing mechanism: providers received predictable payments for delivering high-quality training while a meaningful share of funding depended on verifiable progression into work, strengthening incentives to support participants through the transition into employment.

Programme documentation explains why the results-based share was not set higher. It highlights risks such as employment "churn" in low-skill jobs – cycles of short employment spells followed by returns to unemployment – which would not be well addressed by a payment model based overwhelmingly on employment outcomes. Instead, the report argues for maintaining a combined focus on certified, transferable skills and employment outcomes with the ratio calibrated to participant profiles and labour market conditions (SOLAS, 2016[23]).

Finally, Momentum illustrates how data linkages and verification can underpin results-based payments. The evaluation stresses that employment outcomes were verifiable because programme systems maintained a clear record linking participants to providers and employers, and because outcomes reported by providers were verified for payment and subsequently confirmed through administrative checks. Despite positive reviews of its design and results (SOLAS, 2016[23]), the Momentum programme was allowed to expire in 2017 as economic needs and labour market conditions changed.

Sources: Government of Ireland (2023[19]), *Employment Inclusion Skills and Training (EIST) Programme 2021-2027: Evaluation Plan*, https://eufunds.ie/wp-content/uploads/2023/11/EIST-Evaluation-Plan-Final.pdf (accessed on 12 May, 2026); Indecon (2020[20]), *Evaluation of Specific Skills Training*, https://a.storyblok.com/f/70398/x/9aa70231b7/sst-independent-evaluation-_2020_indecon.pdf (accessed on 12 May, 2026); Kelly et al., (2025[21]), *Roadmap for an evaluation of the National Training Fund*, Economic & Social Research Institute (ESRI) Research Serieshttps://www.esri.ie/system/files/publications/RS221.pdf (accessed on 12 May, 2026); SOLAS (2014[22]), *Programme Evaluation: Momentum Programme for Projects under the Labour Market Education and Training Fund,* Report 5: Final report, https://www.solas.ie/f/70398/x/4981308c21/momentum_evalutation_report.pdf (accessed on 12 May, 2026); SOLAS (2016[23]), Momentum 2 Evaluation 2015-16, *Momentum Programme*, https://www.solas.ie/f/70398/x/b59a926e1c/645-momentum-2016-report-r19.pdf (accessed on 12 May, 2026); OECD (2020[24]), "Education Policy Outlook in Ireland", *OECD Education Policy Perspectives*, No. 18, OECD Publishing, Paris, https://doi.org/10.1787/978e377b-en.

The following paragraphs discuss three main design considerations for including an FNLC approach in Programme "Education".

Use a small, balanced set of indicators anchored in the theory of change

First, indicators that influence funding should reflect the main steps in the ToC:

- Outputs – what was delivered (e.g. number of teachers trained, mentoring sessions held).
- Intermediate outcomes – whether key mechanisms are moving (e.g. teaching practices, use of new resources, attendance of target groups).
- Final outcomes/Results – what changed for learners or institutions (e.g. completion, dropout, achievement, or employment).

Combine Simplified Cost Option with modest result-based top-ups

International guidance on SCOs and FNLC shows that result based payments are usually layered on top of cost-based funding rather than replacing it. In practice, this can be done in two main ways, depending on whether the policy priority is budget-neutrality or risk minimisation for beneficiaries:

- **Option 1. Use a partial base payment and a result-based top-up within the same overall funding envelope.** Under this model, most expected costs are reimbursed through SCO, while a small share of funding is paid only if agreed results are achieved. For example, the Executive

Agency could reimburse 80–95% of the expected cost through SCO and reserve the remaining share for a result-based top-up. The payment levels can be calibrated so that the expected average cost remains close to that under pure SCO. For example, if around half of projects are expected to achieve the target, a 90% base payment and a 20% bonus would be budget-neutral on average. This type of calibration is explicitly discussed in European FNLC guidance (see Table 19 in European Commission (2021[15])). This option may be most suitable if budget-neutrality relative to pure SCO is an important objective.

- **Option 2. Maintain full SCO reimbursement and add a separate result-based bonus.** Under this model, the Executive Agency would continue to reimburse agreed costs in full through SCO and would pay an additional bonus if agreed results are achieved. This removes any financial risk for beneficiaries and may therefore be more suitable where protecting beneficiary liquidity is important, where outcomes are observed only with a delay, or where outcomes may be affected by external factors beyond the beneficiary's control.

Both approaches are frequently used in practice. Under the European Commission's Aim-Learn-Master-Achieve initiative for integrating disadvantaged youths (European Commission, 2024[25]), top-ups for successful participants are paid in addition to the basic unit cost per participant. In other systems, a larger share of total payment is linked to outcomes, including examples from Sweden and the United Kingdom in contracted employment and skills services (Langenbucher and Vodopivec, 2022[26]).

Both approaches keep most funding predictable and tied to activities under the beneficiary's direct control while also ensuring that a small share is explicitly dependent on results. World Bank experience with results-based financing in education also emphasises that only a limited fraction of funding should be performance-linked, and that this share should be carefully calibrated based on risk and capacity (World Bank, 2017[17]; 2021[18]).

Cohort- or population outcomes versus participant-based outcomes

ESF+ common indicators are, by design, almost entirely based on groups of participants. These describe outcomes for people who choose to take part in a particular activity, such as counselling, mentoring, or mobility programmes. Such indicators are well suited for demonstrating the scale and reach of ESF+ support and for feeding EU-level monitoring. At the same time, participant indicators provide limited information about the true impact of an activity on outcomes because participants are typically self-selected and can differ from non-participants in important ways (e.g. more motivated, better informed, etc.). As a result, there is no obvious comparison value for the results achieved by participants, and thus it is difficult to know if the programme made a difference.

For programme- and project-level management, the Executive Agency may thus wish to complement participant-based indicators with a small number of **cohort or population indicators.** These describe outcomes for a whole reference group, such as all students in a school or a degree programme, or all graduates in a field. Cohort or population indicators may not be entirely under the beneficiary's control, but they provide a more meaningful picture of whether the overall situation of the target group is improving in line with the policy goals of Programme "Education". While confounds remain possible, observing improvement in an entire reference group after the introduction of an activity is more suggestive that the activity had a positive impact than the performance of a group of self-selected participants (see Chapter 2, Box 2.1).

A practical way to decide between the use of the two indicator types or decide how to combine types therefore depends on whether reforms are designed to affect entire groups or instead focus on individuals who deliberately choose to participate. Note that, when participation is mandatory and enforced for an entire group, cohort- and participant-based indicators are identical:

- Cohort or population-based FNLC – for reforms that affect whole schools, programmes or regions. Examples include upskilling the teacher workforce, anti-bullying campaigns, curricular reforms in

VET, or competence-based reforms in higher education. In these cases, it is natural to condition a small bonus on cohort-level outcomes such as school-wide dropout rates, average achievement for target groups, or completion rates. World Bank work on results-based financing at sub-national level explicitly recommends using such population indicators to encourage system-level improvements (World Bank, 2021[18]).

- Participant-level FNLC – for services where individuals (typically students or parents) opt in, such as counselling, mentoring, or job-search assistance. Here, result-based payments can be tied to the outcomes of participants (e.g. whether they remain in education, gain a qualification, or find and keep a job).

Box 3.2 demonstrates two short, stylised examples to illustrate how this distinction could work in practice. In both cases, ESF+ rules on FNLC require that the result can be precisely defined, credibly measured, and independently verified, but they do not prescribe whether indicators must be participant- or cohort-based.

Box 3.2. Financing Not Linked to Costs: Simple illustration examples

The following examples demonstrate how either cohort-based or participant-based indicators could be used in Financing Not Linked to Costs (FNLC), depending on whether reforms target larger groups (e.g. all students in a school or grade) or focus on individuals who choose to participate.

Example A – Cohort-based bonus for school-wide dropout reduction:

- *Base payment*: schools receive 100% of the simplified cost option rate for running an evidence-based dropout prevention programme (e.g. mentoring, parental outreach, targeted support).
- *Outcome indicator:* "share of students from vulnerable groups in Grades 5–8 who are still enrolled at the end of the school year", measured from EUMIS data for the whole school.
- *Result-based top-up*: schools receive an extra 20% if the dropout rate of vulnerable students falls below a pre-defined target floor (e.g. a reduction of at least X percentage points compared to the three-year baseline).
- *Rationale*: this encourages schools to focus on the quality and targeting of support not only on enrolling participants. It also uses a cohort-level indicator that captures effects on all relevant students not only those who took part in a specific project.

Example B – Participant-level bonus for employment after higher education:

- *Base payment*: Higher education institutions receive 100% of the simplified cost option rate for providing career guidance, internships, and placement support to students in selected programmes.
- *Outcome indicator*: "share of graduates from the supported programmes who are employed or in further education six months after graduation", measured using linked administrative data (e.g. public employment or social security registers).
- *Result-based top-up*: providers receive an additional 15-20% if this employment or continuation rate meets or exceeds an agreed target band (partial payments are possible depending on reaching target floor, target, ambitious target).
- *Rationale*: this creates a direct incentive to improve the labour-market relevance of support while keeping most funding on a predictable cost basis.

These examples are consistent with international practice, where result-based bonuses typically cover a modest share of total financing and are tied to a small number of clearly defined indicators (World Bank, 2021[18]).

Source: World Bank (2021[18]), *Results-Based Financing in Education for Sub-national Government and Schol Administrators: A Conceptual Framework and Practical Recommendations*, https://documents1.worldbank.org/curated/en/182581644936918794/pdf/Results-Based-Financing-in-Education-for-Sub-national-Government-and-School-Administrators-A-Conceptual-Framework-and-Practical-Recommendations.pdf (accessed on 15 May, 2026).

Beyond these specific design considerations, any use of FNLC in Programme "Education" would need to be calibrated carefully so that result-based elements strengthen attention to quality and learning while limiting administrative burden and unintended behavioural responses. Box 3.3 summarises the main safeguards.

Box 3.3. Results-based finance: Getting incentives right and managing risks

Experience from results-based financing in Europe and elsewhere suggests that result-based elements need to be designed carefully, so that they strengthen attention to meaningful results and programme quality without creating unnecessary administrative pressure, unintended behavioural responses or perverse incentives (World Bank, 2018[27]). In the context of Programme "Education", this means framing results-based elements positively as rewards for performance, keeping their share modest, limiting the number of indicators, and relying on indicators that can be verified independently. This is consistent with European Commission guidance, which cautions against overly mechanistic scoring (European Commission, 2017[28]), and with World Bank guidance, which emphasises careful calibration of performance-linked payments and credible verification arrangements (World Bank, 2017[17]; 2018[27]; 2021[18]). In practice, safeguards could include:

- **Keeping the bonus modest and framing it positively.** A small bonus for strong results is more likely to be seen as recognition for excellent performance than as a sanction. This may help reinforce the existing intrinsic motivation of providers while still drawing greater attention to quality and results.
- **Using a small number of indicators and avoiding overly mechanical scoring rules.** This can help ensure that beneficiaries focus on meaningful progress rather than on narrow compliance with multiple individual targets (i.e. "box-ticking behaviour").
- **Basing payments on administrative or other independently verifiable data wherever possible.** This can reduce misreporting risks and help maintain a clear and proportionate audit trail.
- **Clearly defining how dropouts and missing data are treated in outcome indicators.** For example, non-respondents or participants who leave the programme early may need to be assigned a zero value on the relevant indicator. This can reduce incentives to improve measured performance by excluding participants who are harder to support.
- **Include indicators that reflect progress among participants who start further behind.** If payment depends only on average results, attention may shift towards participants who are easiest to support while those with greater needs receive less support. This can be mitigated by measuring progress from each participant's baseline, or by using indicators that reflect results across the full performance distribution.

- **Piloting the approach in a small number of procedures first and reviewing behaviour after one full cycle.** This would allow the Executive Agency to test whether the indicators, payment levels and verification arrangements are working as intended before any wider rollout.

If designed in this way, results-based finance can shift a small but meaningful part of incentives towards what changes for learners and institutions, while keeping the system manageable for both the Executive Agency and beneficiaries.

Source: World Bank (2018[27]), *Result-Based Financing in Education: Learning from What Works*, https://documents1.worldbank.org/curated/en/915061548222619389/pdf/Results-Based-Financing-in-Education-Learning-from-What-Works.pdf (accessed on 15 May, 2026); European Commission (2017[28]), *Budget Support Guidelines*, Tools and Methods Series Guidelines N°7, https://international-partnerships.ec.europa.eu/system/files/2019-09/budget-support-guidelines-2017_en.pdf (accessed on 13 May, 2026); World Bank (2017[17]), *Results-Based Financing in Education: Financing Results to Strengthen Systems,* https://documents1.worldbank.org/curated/en/715791489054110215/pdf/113265-REVISED-RBF-Approach-Final-Digital-Apr21.pdf (accessed on 15 May, 2026); World Bank (2021[18]), *Results-Based Financing in Education for Sub-national Government and Schol Administrators: A Conceptual Framework and Practical Recommendations*, https://documents1.worldbank.org/curated/en/182581644936918794/pdf/Results-Based-Financing-in-Education-for-Sub-national-Government-and-School-Administrators-A-Conceptual-Framework-and-Practical-Recommendations.pdf (accessed on 15 May, 2026).

References

Andrews, M., P. Lant and W. Michael (2012), "Escaping Capability Traps through Problem-Driven Iterative Adaptation (PDIA)", *Working Paper 299*, https://www.cgdev.org/publication/escaping-capability-traps-through-problem-driven-iterative-adaptation-pdia-working-paper (accessed on 22 May 2026). [8]

Department of Further and Higher Education Reserch, I. (ed.) (2023), *Employment Inclusion Skills and Training (EIST) Programme 2021-2027: Evaluation Plan*, https://eufunds.ie/wp-content/uploads/2023/11/EIST-Evaluation-Plan-Final.pdf (accessed on 12 May 2026). [19]

EAPE (2022), *Methodology for defining the indicator framework*, https://sf.mon.bg/?go=page&pageId=364 (accessed on 22 May 2026). [6]

Education, T. (ed.) (2025), *Theory of Change Builder: Guidance and resources*, https://taso.org.uk/libraryitem/theory-of-change-builder/ (accessed on 23 December 2025). [4]

European Commission (2025), *Study on the uptake of Simplified Cost Options (SCO) and Financing Not Linked to Costs (FNLC) for the Common Provisions Regulation (CPR) Funds in the 2014-2020 and 2021-2027 programming periods*, Publications Office of the European Union, https://ec.europa.eu/regional_policy/information-sources/publications/studies/2025/study-on-the-uptake-of-simplified-cost-options-sco-and-financing-not-linked-to-costs-fnlc-for-the-common-provisions-regulation-cpr-funds-in-the-2014-2020-and-2021-2027-prog (accessed on 22 May 2026). [16]

European Commission (2024), *ALMA: Aim-Learn-Master-Achieve: Active inclusion initiative for integrating disadvantaged young people not in education, employment or training (NEETs) through mobility - Handbook*, https://european-social-fund-plus.ec.europa.eu/system/files/2024-02/ALMA%20Handbook.pdf (accessed on 12 May 2026). [25]

European Commission (2021), *'Simplified cost options' and 'Financing not linked to costs' in the area of social inclusion and youth - A study complementing the ESF+ impact assessment: final report*, Publications Office of the European Union, https://data.europa.eu/doi/10.2767/78485 (accessed on 22 May 2026). [15]

European Commission (2021), *ESF+ Common indicators index: Working document*, https://sfc.ec.europa.eu/system/files/documents/documents/toolbox-october-2021_0.pdf (accessed on 23 December 2025). [5]

European Commission (2017), "Budget Support Guidelines", *Tools and Methods Series Guidelines N°7*, https://international-partnerships.ec.europa.eu/system/files/2019-09/budget-support-guidelines-2017_en.pdf (accessed on 13 May 2026). [28]

Harvard Kennedy School (2025), "Designed to guide you through the process of solving complex problems while building capability", *PDIA Toolkit*, https://bsc.hks.harvard.edu/tools/toolkit/ (accessed on 26 December 2025). [14]

Indecon (2020), *Evaluation of Specific Skills Training*, https://a.storyblok.com/f/70398/x/9aa70231b7/sst-independent-evaluation-_2020_indecon.pdf (accessed on 12 May 2026). [20]

Kelly, E. et al. (2025), "Roadmap for an evaluation of the National Training Fund", *Economic & Social Research Institute (ESRI) Research Series*, https://www.esri.ie/system/files/publications/RS221.pdf (accessed on 12 May 2026). [21]

Khattri, N. and D. Roberts (2012), *Designing a Results Framework for Achieving Results: A How-to-guide (English)*, Washington D.C.: World Bank Group, https://documents.worldbank.org/en/publication/documents-reports/documentdetail/331541563854787772 (accessed on 22 May 2026). [3]

Langenbucher, K. and M. Vodopivec (2022), *Paying for results: Contracting out employment services through outcome-based payment schemes in OECD countries*, OECD Publishing, Paris,, https://doi.org/10.1787/c6392a59-en. [26]

OECD (2026), *Implementing inclusive education projects in Bulgaria: Lessons from an evaluation of Bright Start and Success for You*, OECD Publishing, Paris,, https://doi.org/10.1787/030e0717-en. [13]

OECD (2025), *Implementation Toolkit for the OECD Recommendation on Public Policy Evaluation*, OECD Publishing, Paris, https://doi.org/10.1787/77faa4fe-en. [9]

OECD (2025), *Modernisation of Higher Education Institutions Procedure in Bulgaria: Interim Evaluation and Lessons for Educational Programming*, OECD Publishing, Paris, https://doi.org/10.1787/cddcd19e-en. [11]

OECD (2023), *Who Really Cares about Using Education Research in Policy and Practice?: Developing a Culture of Research*, OECD Publishing, Paris, https://doi.org/10.1787/bc641427-en. [12]

OECD (2022), *Who Cares about Using Education Research in Policy and Practice?: Strengthening Research Engagement*, OECD Publishing, Paris, https://doi.org/10.1787/d7ff793d-en. [10]

OECD (2021), *Applying Evaluation Criteria Thoughtfully*, OECD Publishing, Paris, https://doi.org/10.1787/543e84ed-en. [1]

OECD (2020), *Education Policy Outlook in Ireland*, OECD, Publishing, Paris, https://www.oecd.org/content/dam/oecd/en/publications/reports/2020/12/education-policy-outlook-in-ireland_a2b63d91/978e377b-en.pdf (accessed on 22 May 2026). [24]

SOLAS (2016), "Momentum 2 Evaluation 2015-16", *Momentum Programme*, https://www.solas.ie/f/70398/x/b59a926e1c/645-momentum-2016-report-r19.pdf (accessed on 12 May 2026). [23]

SOLAS (2014), "Programme Evaluation: Momentum Programme for Projects under the Labour Market Education and Training Fund", *Report 5: Final report*, https://www.solas.ie/f/70398/x/4981308c21/momentum_evalutation_report.pdf (accessed on 12 May 2026). [22]

Vogel and Isabel (2012), *DFID research: Review of the use of 'Theory of Change' in International development*, https://www.gov.uk/government/news/dfid-research-review-of-the-use-of-theory-of-change-in-international-development (accessed on 22 December 2025). [2]

World Bank (2021), *Results-Based Financing in Education for Sub-national Government and Schol Administrators: A Conceptual Framework and Practical Recommendations*, https://documents1.worldbank.org/curated/en/182581644936918794/pdf/Results-Based-Financing-in-Education-for-Sub-national-Government-and-School-Administrators-A-Conceptual-Framework-and-Practical-Recommendations.pdf (accessed on 15 April 2026). [18]

World Bank (2018), *Result-Based Financing in Education: Learning from What Works*, https://documents1.worldbank.org/curated/en/915061548222619389/pdf/Results-Based-Financing-in-Education-Learning-from-What-Works.pdf (accessed on 13 May 2026). [27]

World Bank (2017), *Results-Based Financing in Education: Financing Results to Strengthen Systems*, https://documents1.worldbank.org/curated/en/715791489054110215/pdf/113265-REVISED-RBF-Approach-Final-Digital-Apr21.pdf (accessed on 15 May 2026). [17]

Zall Kusek, J. and R. Rist (2004), *Ten Steps to a Results-Based Monitoring and Evaluation System : A Handbook for Development Practitioners*, World Bank, https://openknowledge.worldbank.org/entities/publication/84e1b3ba-74a2-5a19-a68d-a4b9632b74ad (accessed on 22 May 2026). [7]

4 Strengthening system-level data use

This chapter broadens the focus beyond Programme “Education” 2021-2027, looking at the education system more broadly. It considers how Bulgaria can strengthen evaluation processes and data use in a way that remains fully compliant with data-protection rules and how this in turn might support the evaluation of both Programme “Education” and other education programmes and initiatives, enabling them to better inform future programme design. It discusses ways to design indicators with practical data access requirements in mind from the outset, to plan robust and transparent data-sharing arrangements between institutions and to make greater use of existing administrative data for evaluation.

4.1. Current status and challenges of system-level data use

To design meaningful indicators and make effective use of administrative data, it is essential to understand the current status of system-level data in Bulgaria, including the practical challenges inherent in accessing, linking and using these data across institutions. A stronger focus on outcomes, including the mechanisms and intermediate outcomes that underpin them makes the availability of the right outcome data becomes essential. If Programme “Education” increasingly wishes to understand whether reforms are improving learning, inclusion or labour-market outcomes, it needs data that describe those outcomes for the relevant groups, such as all students in a grade or programme, or all graduates from a vocational education and training (VET) school. Activity and participant data remain necessary, but they are not sufficient to understand whether the programme is achieving its goals.

At present, much of the information used in monitoring comes from beneficiaries’ own reporting, verified through ex-post audits. This is appropriate for outputs that are under the direct control of beneficiaries, such as the number of training sessions delivered, or participants reached. However, many of the outcomes that matter for Programme “Education” are not directly visible to beneficiaries. Examples include students’ performance in national assessments, completion and dropout rates over several years, or graduates’ employment and earnings after leaving education. These outcomes are usually recorded in administrative systems, such as student registers, examination databases, social security records or employment registers, not in project reports.

For these reasons, administrative data held by public bodies become a key resource. The use of aggregate statistics may be tempting, given their relatively easy availability. Yet in many cases, aggregate statistics alone are not enough. Programme indicators may be defined at a different level of aggregation than the available statistics. For instance, a regional employment rate is not sufficient if the relevant indicator is, *“*employment rate of graduates from VET school X six months after graduation”. Verifying such indicators typically requires matching records across datasets at the level of individual learners or graduates, using identifiers that allow education records to be linked to employment, social security or other registers. Individual-level data are also needed when there is a wish to construct treated groups and comparison groups that have faced similar conditions outside of the intervention of interest. Finally, individual records greatly facilitate subgroup comparisons across characteristics of interest.

Accessing individual data understandably raises questions about privacy and data protection. However, European experience shows that secure use of administrative data for monitoring and evaluation is both

possible and lawful when it is clearly framed as a task carried out in the public interest and when strong safeguards are in place. Under the General Data Protection Regulation (GDPR), public authorities may process personal data where this is necessary for performing a task in the public interest set out in law (Article 6(1)(e)), and they may process special category data for research and statistical purposes when appropriate safeguards are applied (Article 9(2)(j) and Article 89). Recent European Union (EU) guidance for European Social Fund (ESF) and European Social Fund Plus (ESF+) programmes confirms that administrative data can be reused for monitoring and evaluation when roles, safeguards and responsibilities are clearly defined (European Commission, 2023[1]).

The main challenge is therefore not whether data can be used, but how to plan and organise this use well. To make best use of existing data while respecting privacy, indicators need to be designed with data sources and access in mind from the beginning. Institutions need clear agreements on who holds which data, how they can be securely linked for specific evaluation purposes, and how outputs are returned in an aggregated, de-identified form. Done well, this can reduce the reporting burden on beneficiaries, avoid repeated ad hoc data requests, and give Programme "Education" a stronger evidence base for deciding where and how to invest limited resources.

4.2. Designing Programme "Education" indicators with a data access plan from the start

Programme "Education" already has a strong starting point in its indicator work. The 2022 methodology (EAPE, 2022[2]) and the associated indicator factsheets clearly define each indicator, its numerator and denominator, and the basic data collection methodology. This is very good practice and already aligns with ESF+ guidance. As the focus of indicators turns further towards impacts on cohorts and populations, the next step could be to make data access and linkage an explicit part of these fact sheets.

A practical way to do this is for the Executive Agency to expand the "data collection methodology" section in each fact sheet so that, when an outcome indicator is defined at programme or project level, a short "data access plan" is recorded alongside it. For each such indicator, the fact sheet could explicitly state:

- **What data are needed.** Specify whether the indicator refers to a whole reference group (e.g. all Grade 8 students in a school, all VET graduates in a year) or only to participants, and whether it requires individual records or only aggregated figures (and if so, at what level). It should also note the relevant timing (e.g. at the end of the school year, 6 or 12 months after graduation).
- **Who collects and holds the data and in what format.** Identify the primary data holder, such as the Ministry of Education and Science (hereafter, Ministry), the National Statistical Institute (NSI), the Employment Service, the National Revenue Agency, or social security institutions. Where several bodies hold related data, this should be recorded.
- **How targets will be verified and by whom.** Indicate whether verification will rely on direct access by Executive Agency Programme "Education" (hereafter, Executive Agency) or Ministry staff to an administrative register, on a formal linkage between the Managing Authority and a register holder, or on a trusted third party (for example, the NSI) that performs the linkage and returns anonymised aggregates or flags. Questions of access and data protection are explored in the remainder of this section.

For many indicators, the necessary data already exist in Bulgarian systems. Making these points explicit in the fact sheet would avoid later uncertainty about who needs to provide what and would reduce the need for repeated ad hoc data requests once projects are underway.

Fact sheets can also be used to anticipate data gaps. Where the necessary data do not currently exist, two additional questions should be explored:

- **If the required data do not currently exist, who is best placed to collect them, and is this feasible at reasonable cost and burden?** For example, if a secondary school wants to track student satisfaction with a new course or campaign, a short end-of-term survey may be realistic.
- **If the required data cannot feasibly be collected, what is a practical and credible alternative indicator that serves a similar purpose?** For instance, if detailed employment status is not available in a timely way, paid social security contributions could be used as a proxy for being in work. If detailed test-score distributions are not available, the share of students reaching a minimum proficiency level in national assessments may be a workable alternative.

Embedding answers to these questions into the existing indicator fact sheets would not fundamentally change the current framework. It would simply ensure that, whenever a new outcome indicator is proposed, the availability, ownership and access path for the underlying data are considered from the beginning. This reduces subsequent delays, supports the use of administrative data for verification, and helps Programme "Education" move from "what should we measure?" to "what can we reliably measure, with which data, and through which institutional route?".

4.3. Anticipating challenges: individual-level data and data protection

There are widespread concerns in Bulgaria that the use of individual-level data is illegal. However, these concerns can be addressed: GDPR and national frameworks are designed to enable responsible, public-interest data access for exactly the kinds of evaluation questions that Programme "Education" needs to answer.

Interviews for this report – including with experts at NSI and the Institute of Education – suggest a widespread concern that individual-level data cannot lawfully be shared or reused. Once a dataset contains personal identifiers, such as names, personal numbers or dates of birth, many officials feel it must remain siloed within the institution that collected it. This perception is understandable given the importance of data protection.

However, under the GDPR, public bodies may process personal data when this is necessary to perform a task carried out in the public interest or in the exercise of official authority, provided that appropriate safeguards are in place (Article 6(1)(e)) (European Parliament and Council, 2016[3]; ISO, 2025[4]). For research, statistics and archiving in the public interest, the GDPR allows processing subject to, "appropriate safeguards", including technical and organisational measures and respect for data minimisation (Article 89(1)) (European Parliament and Council, 2016[3]). In practice, this means that individual-level data can be used for well-defined public interest purposes, such as evaluating education reforms as long as the use is necessary, proportionate, and properly safeguarded.

On this legal basis, many EU and OECD countries have built secure services that give controlled access to individual-level data for statistical and research purposes while fully respecting GDPR and national law. For example:

- **Inter-administration data exchange.** In France, administrative registers from different public bodies (e.g. employment, unemployment and training records) are linked for statistical and evaluation purposes. The FORCE dataset, for instance, combines information on participation in training with subsequent labour-market outcomes to study the effectiveness of programmes. Poland's graduate tracking system provides another example, linking higher education records with social insurance data to monitor graduates' labour market outcomes under a secure and anonymised system (see Box 4.1). Such linkages are typically performed by a trusted body (e.g. the NSI in Bulgaria) under a clear legal mandate and strict, formalised confidentiality rules.
- **Secure access for external researchers.** Several countries operate "safe rooms" or secure remote environments where approved external researchers can analyse de-identified microdata

but cannot remove raw records. The Secure Data Access Centre (*Centre d'accès sécurisé aux données)* in France provides secure access to a wide range of sensitive datasets for scientific research, under GDPR-certified infrastructure (CASD, 2025[5]). The United Kingdom's Office for National Statistics' Secure Research Service and Ireland's Central Statistics Office's Researcher Data Portal follow similar "Five Safes" principles: only accredited users are allowed in; access is via locked-down systems; microdata never leave the secure server; and all outputs are checked before release (ONS, 2022[6]; CSO, 2024[7]). At EU level, Eurostat provides access to confidential microdata for recognised research entities under Regulation (EC) No 223/2009, again with strict safeguards for privacy and statistical confidentiality (Eurostat, 2025[8]).

Box 4.1. Monitoring and evaluation of European Social Fund Plus projects in Poland

Like Bulgaria, Poland operates under the common European Commission requirements for monitoring and evaluation of European Social Fund Plus (ESF+) programmes. Poland therefore prepares comprehensive evaluation plans for each programme co-financed by ESF+, which define the scope of studies, outline possible methodological approaches and specify related responsibilities. The Managing Authority in Poland also issues evaluation guidelines that apply to all programmes implemented with funds from the European Union.

Poland's ESF+ monitoring and evaluation system is characterised by a comparatively strong evaluation infrastructure that supports learning and follow-through. First, Poland has invested in integrated digital systems for programme administration and monitoring. The Centralised IT System 2021 was designed to support the electronic exchange of information between beneficiaries and programme authorities, including the routine handling of project documentation and implementation data. Second, Poland maintains a dedicated national evaluation portal (ewaluacja.gov.pl) where evaluation reports are made available, and an accompanying spreadsheet-based archive of evaluation studies (Ministry of Funds and Regional Policy, n.d.[9]). Third, Poland operates a Recommendation Implementation System that records recommendations from evaluation studies and supports tracking their implementation status, which provides a concrete mechanism for closing feedback loops and moving from evidence to action (Ministry of Funds and Regional Policy, 2026[10]).

Use of theories of change and intervention logic. In Poland, the language of intervention logic and theory of change is routinely used in evaluation planning and reporting, including in regional programme evaluations. For example, an evaluation of pre-school interventions under the Mazowieckie regional programme explicitly frames the study around the theory of change support (intervention logic) and uses this as a basis for the evaluation questions (EU-CONSULT, 2021[11]). This illustrates a more institutionalised expectation that evaluations should make explicit how activities are intended to lead to outcomes and then test whether the expected pathways hold in practice.

Counterfactual impact evaluation. Poland has also implemented counterfactual impact evaluations for ESF-funded programmes, notably for a Youth Employment Initiative and other labour market interventions (Mado et al., 2021[12]). In education measures under ESF, counterfactual approaches have also been applied, albeit more selectively. A 2019 evaluation of the provision of pre-school access, improved facilities and remedial classes explicitly reports the use of difference-in-differences methodologies (Ministry of Funds and Regional Policy, 2026[13]). In addition, a national evaluation of teacher professional development in the use of information and communication technology in education reports the use of counterfactual analyses to estimate training effects (Ministry of Funds and Regional Policy, n.d.[14]). Looking ahead, Poland is also commissioning work to estimate long-term result indicators for ESF+ participants, signalling an institutional push to move from process tracking towards outcome measurement that can support stronger impact evaluation where feasible.

Linked administrative data for outcome tracking. Beyond programme monitoring systems, Poland has built ways to link administrative data to enable routine outcome tracking of transitions from education to the labour-market. This matters for ESF+ evaluations because it reduces reliance on beneficiary surveys and creates scope for stronger outcome measurement and more credible counterfactual designs, when appropriate. Poland's Graduate Tracking System (ELA) is a nationwide system that combines higher-education records with labour-market information from the Social Insurance Institution (ZUS) (POL-on, n.d.[15]; ELA, n.d.[16]; Bożykowski and Jasiński, 2024[17]). ELA generates a rich set of labour-market outcome indicators for higher education graduates that can be followed over time and disaggregated by field of study, institution, and cohort (ELA, n.d.[18]). Observed outcomes include employment incidence, unemployment risk, earnings outcomes, job stability, time from graduation to employment, working while studying, geographic mobility, dropout dynamics, and others (ELA, 2025[19]).

ELA illustrates how individual-level administrative data can be used to move beyond aggregate statistics towards cohort-based outcome tracking. ELA's value stems from a routinised data pipeline that enables individual-level linkages for all higher education graduates in Poland while returning data outputs in a secure form. Graduate records from the digital system for collecting data in Polish higher education are transferred to ZUS, which enriches them with social-insurance based labour-market outcomes, then pseudonymises the data before it returns them for analysis and reporting. The data flow and timelines are set out in law, including specific provisions under Article 352 of the 2018 Law on Higher Education and Science (Bożykowski and Jasiński, 2024[17]).

ELA is an example of how a clear legal mandate can be operationalised to enable data use for the public interest. Even though ELA is not an ESF+ tool per se, it illustrates how indicators can be defined at the level of cohorts and programmes and how data linkages can enable subgroup analysis and quantitative impact evaluation designs when appropriate. The ELA system continues to evolve: official documentation notes ongoing work to extend analytical scope, including deeper integration with tax data and links to earlier stages of education, which would further strengthen the system's ability to follow individuals across education and employment pathways.

Source: Ministry of Funds and Regional Policy (n.d.[9]), Evaluation study results, *Evaluation Service,* https://www.ewaluacja.gov.pl/strony/badania-i-analizy/wyniki-badan-ewaluacyjnych/ (accessed on 15 May, 2026); Ministry of Funds and Regional Policy (2026[10]), *Recommendation Management System (SWR) Base*, https://baza.ewaluacja.gov.pl/ (accessed on 15 May, 2026); EU-CONSULT (2021[11]), *Evaluation of measures undertaken in the area of childcare for children under 3 and educational-care services for preschool children under the Regional Operational Programme for Mazowieckie Voivodeship 2014–2020: Final report.,* Mazowieckie Voivodeship, https://www.funduszedlamazowsza.eu/wp-content/uploads/2021/08/rk-rpo-wm-wychowanie-przedszkolne-14.05.2021.pdf (accessed on 15 May, 2026); Mado et al., (2021[12]), "What works for whom? Youth labour market policy in Poland", IZA Discussion Paper No. 14793, https://docs.iza.org/dp14793.pdf (accessed on 15 May, 2026); Ministry of Funds and Regional Policy (2026[13]), Evaluation of measures undertaken in support of education: Final report, *Evaluation Service,* https://www.ewaluacja.gov.pl/media/77466/RK-Ewaluacjadzialanpodejmowanychnarzeczedukacji.pdf (accessed on 15 May, 2026); Ministry of Funds and Regional Policy (n.d.[14]), Education under the European Social Fund: Third partial report, *Evaluation Service,* https://www.ewaluacja.gov.pl/media/105765/Edukacja_EFS_III_raport_czastkowy_PUBLIKACJA.pdf (accessed on 15 May, 2026); POL-on (n.d.[15]), *POL-on network*, https://polon.nauka.gov.pl/en/polon-network/ (accessed on 15 May, 2026); ELA (n.d.[16]), *National monitoring system of the economic trajectories of higher education graduates,* https://ela.nauka.gov.pl/pl/home (accessed on 15 May, 2026); Bożykowski and Jasiński (2024[17]), *Polish Graduate Tracking System: The analytical power of administrative data: Presentation*, https://skilltracking.highereducation.ac.cy/wp-content/uploads/2024/05/Polish-Graduate-Tracking-System-the-analytical-power-of-administrative-datav2.pdf (accessed on 15 May, 2026); ELA (n.d.[18]), ELA: Graduate tracking system, https://ela.nauka.gov.pl/en (accessed on 15 May, 2026); ELA (2025[19]), ELA system – autumn release of the 10th edition, https://ela.nauka.gov.pl/pl/labor-market/system_ELA_jesienna_odslona_X_edycji (accessed on 15 May, 2026).

These examples show that using individual-level data for public-interest statistics and evaluation is not an exception, but standard practice in many OECD systems. The crucial condition is thus not to avoid individual-level data, but to define clear public-interest purposes, document responsibilities, and apply robust technical and organisational safeguards. For Programme "Education", this means that

administrative identifiers may be used within a secure environment to link, for example, graduate lists for VET programmes to employment or social security registers, but any results shared outside that secure environment would be aggregate indicators, such as employment rates by school or region, with no way to identify individual learners.

4.4. Creating a simple, secure route to link and use data across institutions

Bulgaria already has much of the legal and technical infrastructure needed for safe, public-interest use of individual-level data. At the centre of this infrastructure is the NSI. The Statistics Act gives the NSI a clear mandate to obtain data from other public bodies for statistical purposes, including individual-level records. Article 20(6) specifies that central bodies, local government authorities and other organisations maintaining registers are obliged to provide the NSI with the data it requires - including individual data - for statistical surveys, subject to strict confidentiality rules (NSI, 2023[20]). In parallel, Bulgaria's e-government and interoperability framework has been designed to allow lawful data exchange between public administrations. A central component of this is RegiX, the national inter-register exchange software, which connects dozens of administrative registers and allows automated queries between them while respecting interoperability and data-protection requirements (Bower, 2023[21]). Recent OECD work on Bulgaria's National Employment Agency shows how RegiX is already used in practice to pull information such as previous employment records directly from connected registers into administrative systems (OECD, 2025[22]).

Taken together, this means Bulgaria does not need to build a data-sharing system from scratch. The legal basis for public-interest data use exists, the NSI already has the authority to receive and process identifiable records for statistical purposes, and RegiX provides a technical backbone for secure exchanges between registers. Where there is still work to do is in turning these components into a simple, well-understood and consistently applied route for using linked data to support evaluation and learning in education. At present, education stakeholders report that they experience individual-level data as inaccessible - or are unaware it exists - even when the law would allow controlled access. This suggests a need for clearer operational rules, standard routines for requesting and linking data, and more proactive communication.

A single secure data service, hosted by a public body

One practical step would be for the Bulgarian authorities to establish a single secure data service, hosted by a public body such as the NSI, with a clear mandate to support evidence-informed policy. In this model, the NSI would receive identifiable lists from education authorities (e.g. cohorts of VET or higher education graduates or lists of schools participating in a reform) and then securely link them to existing registers, such as education, employment or social security records. The NSI would then return only aggregated outputs or simple flags (e.g. the share of graduates employed six months after completion, or a yes/no employment status for each anonymous record) so that no personal identifiers leave the NSI environment. This is very similar to the way secure data services operate in other European countries, such as France and the United Kingdom, where researchers and administrations work with de-identified microdata inside a controlled environment and only vetted tables and figures can be exported (Bower, 2023[21]; ONS, 2022[6]).

Such a service could serve two main functions. First, it would simplify inter-administration data exchange for programme monitoring and results-based financing. Instead of negotiating data-sharing agreements for each individual project, the Executive Agency and the Ministry could rely on a standardised process: submit a list of cohorts or institutions, specify the indicators required, and receive a consistent, comparable set of outcome measures across projects and years. Second, it would create a robust platform for evaluation and research. Under strict access rules, approved analysts within government and, where appropriate, external

researchers could use de-identified, linked datasets to study which interventions are most effective, for whom, and under what conditions, without ever seeing names or national identification numbers. This would bring Bulgaria closer to the practice in many high-income countries, where secure research data centres are an important pillar of the evidence base for education and labour-market policy (United Nations, 2022[23]).

For Programme “Education”, the benefits of such a secure route to linked data would be concrete. It would make it feasible to verify cohort-based outcome indicators (e.g. completion, progression, or employment rates for target groups) using administrative registers rather than beneficiary-reported data. It would support fairer designs for Financing Not Linked to Costs because outcome measures could be calculated consistently across schools and higher education institutions. It would also reduce duplication and delays as education authorities would no longer need to submit repeated bilateral requests to each register-holding institution. In short, the existing Bulgarian framework already offers a strong foundation for data-driven evaluation; a next useful step would be to organise it into a clear, user-friendly service that helps Programme “Education” stakeholders turn administrative data into actionable evidence on quality and impact.

References

Bower, L. (2023), *Bulgaria’s Digital Transformation*, AHFE International, USA, https://doi.org/10.54941/ahfe1002953. [21]

Bożykowski, M. and M. Jasiński (2024), *Polish Graduate Tracking System: The analytical power of administrative data*, https://skilltracking.highereducation.ac.cy/wp-content/uploads/2024/05/Polish-Graduate-Tracking-System-the-analytical-power-of-administrative-datav2.pdf (accessed on 15 May 2026). [17]

CASD (2025), *Le CASD – Centre d’accès sécurisé aux données*, https://www.casd.eu/le-centre-dacces-securise-aux-donnees-casd/le-casd/ (accessed on 28 December 2025). [5]

CSO (2024), “Policies - Data for researchers”, *Central Statistics Office, Ireland*, https://www.cso.ie/en/aboutus/lgdp/csodatapolicies/dataforresearchers/policies/ (accessed on 28 December 2025). [7]

EAPE (2022), *Methodology for defining the indicator framework*, https://sf.mon.bg/?go=page&pageId=364 (accessed on 22 May 2026). [2]

ELA (2025), *ELA system – autumn release of the 10th edition*, https://ela.nauka.gov.pl/pl/labor-market/system_ELA_jesienna_odslona_X_edycji (accessed on 15 May 2026). [19]

ELA (n.d.), *ELA: Graduate tracking system*, https://ela.nauka.gov.pl/en (accessed on 15 May 2026). [18]

ELA (n.d.), *National monitoring system of the economic trajectories of higher education graduates*, https://ela.nauka.gov.pl/pl/home (accessed on 15 May 2026). [16]

EU-CONSULT (2021), “Evaluation of measures undertaken in the area of childcare for children under 3 and educational-care services for preschool children under the Regional Operational Programme for Mazowieckie Voivodeship 2014–2020: Final report.”, *Mazowieckie Voivodeship.*, https://www.funduszedlamazowsza.eu/wp-content/uploads/2021/08/rk-rpo-wm-wychowanie-przedszkolne-14.05.2021.pdf (accessed on 15 May 2026). [11]

European Commission (2023), *Smart ways to monitor and evaluate the ESF: How to Gain Access to Administrative Data while Complying with Data Protection Rules - Final Report*, https://op.europa.eu/en/publication-detail/-/publication/343a5ccc-5e87-11ee-9220-01aa75ed71a1/language-en (accessed on 12 May 2026). [1]

European Parliament and Council (2016), "Regulation (EU) 2016/679 (General Data Protection Regulation)", *Official Journal L 110, 4 May 2016*, https://gdpr-info.eu/ (accessed on 22 May 2026). [3]

Eurostat (2025), *Statistical confidentiality and personal data protection*, https://ec.europa.eu/eurostat/web/microdata/statistical-confidentiality-and-personal-data-protection (accessed on 28 December 2025). [8]

ISO (2025), "Definition of "Public Task"", *Information Commissioner's Office guidance on UK GDPR lawful bases*, https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/lawful-basis/a-guide-to-lawful-basis/public-task/ (accessed on 28 December 2025). [4]

Mado, K. et al. (2021), "What works for whom? Youth labour market policy in Poland", *IZA Discussion Paper No. 14793*, https://www.iza.org/publications/dp/14793/what-works-for-whom-youth-labour-market-policy-in-poland (accessed on 15 May 2026). [12]

Ministry of Funds and Regional Policy (2026), "Evaluation of measures undertaken in support of education: Final report", *Evaluation Service*, https://www.ewaluacja.gov.pl/media/77466/RK-Ewaluacjadzialanpodejmowanychnarzeczedukacji.pdf (accessed on 15 May 2026). [13]

Ministry of Funds and Regional Policy (2026), *Recommendation Management System (SWR) Base*, https://baza.ewaluacja.gov.pl/ (accessed on 15 May 2026). [10]

Ministry of Funds and Regional Policy (n.d.), "Education under the European Social Fund: Third partial report", *Evaluation Service*, https://www.ewaluacja.gov.pl/media/105765/Edukacja_EFS_III_raport_czastkowy_PUBLIKACJA.pdf (accessed on 15 May 2026). [14]

Ministry of Funds and Regional Policy (n.d.), "Evaluation study results", *Evaluation Service*, https://www.ewaluacja.gov.pl/strony/badania-i-analizy/wyniki-badan-ewaluacyjnych/ (accessed on 15 May 2026). [9]

NSI (2023), *Statistics Act of the Republic of Bulgaria (amended version)*, https://www.nsi.bg/en/file/25864/StatisticsAct_en.pdf (accessed on 28 December 2025). [20]

OECD (2025), *Optimising Processes and Services at Bulgaria's National Employment Agency*, OECD Publishing, Paris, https://doi.org/10.1787/4e79e9db-en. [22]

ONS (2022), "About the Secure Research Service", *Office for National Statistics*, https://www.ons.gov.uk/aboutus/whatwedo/statistics/requestingstatistics/secureresearchservice/aboutthesecureresearchservice (accessed on 28 December 2025). [6]

POL-on (n.d.), *POL-on network*, https://polon.nauka.gov.pl/en/polon-network/ (accessed on 15 May 2026). [15]

United Nations (2022), *United Nations E-Government Survey 2022: The Future of Digital Government,*, United Nations, https://publicadministration.un.org/egovkb/en-us/reports/un-e-government-survey-2022 (accessed on 22 May 2026). [23]

5 Considering experimental and quasi-experimental evaluation options in procedure design

This chapter explores how choices made during the design of activities and procedures can create opportunities, or constraints, for different types of evaluations. The Executive Agency Programme "Education" (hereafter, Executive Agency) has expressed interest in using some experimental and quasi-experimental methods to strengthen causal inference about specific activities. This interest is well aligned with international practice: experimental and quasi-experimental methods are increasingly used for answering questions of the form "did this intervention cause a change, and by how much?" in policy settings, when used alongside good theories of change and descriptive monitoring.

At the same time, international guidance stresses that such methods are demanding in terms of design and implementation. They typically require clear rules about who receives an intervention and when as well as stable implementation conditions, which can pose logistical, political or ethical difficulties in real-world education settings. They are not appropriate for all evaluation questions, may offer limited qualitative or contextual insight, and cannot replace a broader, multi-method monitoring and evaluation system. They tend to be most useful when there is a clearly defined intervention, a well-specified target group, a small number of priority outcome indicators and realistic options for constructing a comparison or "control" group (HM Treasury, 2020[1]; European Commission, 2021[2]).

For Programme "Education" and future programming, this implies that experimental and quasi-experimental approaches are unlikely to be suitable for evaluating whole procedures that cover several activities and target groups. For such complex procedures, theory-based approaches – including contribution analysis and realist evaluation – are often more appropriate for the overall evaluation, as they focus on how different elements contribute to observed changes in specific contexts. Within such broader evaluations, however, experimental and quasi-experimental methods can still be considered as a targeted tool for specific activities where exposure varies across individuals, classes, or schools, and where meaningful outcome data can be obtained for both those who did and did not receive the intervention.

5.1. Enabling conditions and suitable contexts for different quasi-experimental or experimental designs

This section provides an overview of the main experimental and quasi-experimental designs, focusing on their implications for programme design and data needs, rather than on statistical details. It aims to provide the reader with insights into which methodological approaches are suited to which types of situations. In practice, the use of these methods for evaluation usually depends on a combination of favourable contextual conditions and suitable design choices by the managing authority. The descriptions in this section draw on standard impact-evaluation references and European Social Fund Plus (ESF+) guidance (Gertler et al., 2016[3]; European Commission, 2019[4]; 2020[5]; 2021[2]).

Randomised controlled trials

In a randomised controlled trial, eligible units (e.g. individuals, classes, schools, or small geographical units) are allocated by lottery to a "treatment" group, which receives the intervention, and a "control" group, which either receives it later or not at all. Because the two groups differ only by chance, differences in outcomes can be attributed to the intervention with high confidence. For Programme "Education" and future programming periods, randomised controlled trials are most feasible when:

- There is limited capacity of places and excess demand. For example, a mentoring programme that can serve only a subset of interested students, or a teacher-training course with a fixed number of places.
- It is acceptable to phase the intervention in over time, so that some eligible units receive it later. Note that so-called phase-in designs are limited to observing short-term impacts.
- If access to the intervention is universal among eligible groups, randomised controlled trials can still be feasible if there is scope to randomise encouragement to participate (e.g. additional information, reminder emails, or personalised invitations).
- Outcomes of interest can be measured for both groups (i.e. those receiving the intervention and those not), for example, test scores, attendance or employment outcomes, depending on the context. This condition is necessary in all cases.

Fairness concerns can often be addressed by randomising access only when demand exceeds supply, or by randomising the order in which schools or regions are served (e.g. some schools receive the new support this year, others next year). European Commission peer-review work on experimental evaluations of social inclusion policies describes several such "roll-out" designs as acceptable and politically feasible when transparently communicated (European Commission, 2023[6]).

Regression discontinuity designs

Regression discontinuity designs exploit a sharp rule in the allocation of support. For example, students below a certain test-score threshold might receive additional tutoring, while those above do not. By comparing outcomes for individuals just below and just above the cut-off, evaluators can estimate the effect of the support near that threshold. To make regression discontinuity designs feasible in Programme "Education" and future programming periods:

- Eligibility rules need to be clearly defined by a numeric score, index, or age.
- Nothing else important changes discontinuously at the same cut-off. The only sharp difference at the threshold should be the probability of receiving the intervention.
- The number of cases around the cut-off must be large enough to support analysis.
- High-quality outcome data are required for all eligible individuals, not only those that receive the intervention or participate in the activity.

Regression discontinuity designs can provide impact estimates with strong internal validity when such rules exist and are implemented consistently in practice (European Commission, 2019[4]; 2020[5]).

Difference-in-differences

Difference-in-differences compares the before-after change of a group that receives an intervention with the change over the same time period in a comparison group that does not receive it. For this approach to be informative and feasible under Programme "Education" and future programming periods:

- Outcomes must be measured for both groups (i.e. those receiving the intervention and those not) before the intervention and after it.

- The treated and comparison groups should plausibly follow similar trends in key outcomes of interest over time in the absence of the intervention (the "parallel trends" assumption).
- Administrative or survey data must allow the same groups to be tracked over time.

For example, if some regions introduce a new school-support scheme earlier than others, and student outcomes are measured annually, difference-in-differences can compare changes over time between early and late adopters. Case studies from European Social Fund projects show that such designs are feasible when administrative data contain repeated outcome measures for the relevant groups (European Commission, 2020[5]).

Matching methods

Matching methods compare treated participants with non-participants who are similar in observable characteristics, such as age, prior achievement, socio-economic characteristics or location. They estimate the effect of the intervention by comparing average outcomes between matched groups. While relatively easy to apply, matching has important limitations that should be considered when exploring their application to Programme "Education" or future programming:

- It can only adjust for observed characteristics. Participation is typically influenced by unobserved factors, such as motivation, information, or family support. These factors directly affect outcomes, leading to biased estimates.
- It requires detailed individual-level data for both participants and non-participants.

Matching is usually considered a weaker design and is best complemented by sensitivity checks (Gertler et al., 2016[3]; European Commission, 2019[4]).

Instrumental variables approaches

Instrumental variables approaches use an external factor – an "instrument" – that affects the likelihood of receiving the intervention but does not directly affect the outcome except through that intervention. Examples include distance to the nearest training centre or variation in programme intensity driven by administrative rules.

In practice, suitable instruments are rare and context specific. Instrumental variables approaches typically emerge from detailed knowledge of implementation rather than from simple design choices. For Programme "Education", instrumental variables approaches are unlikely to be a first-line tool, but existing administrative rules or policy changes may sometimes create opportunities for such analyses when carefully assessed (Gertler et al., 2016[3]).

5.2. When might experimental and quasi-experimental methods be feasible in Programme "Education" or in future programming?

A first step is to identify the types of activities for which experimental and quasi-experimental methods have value and are feasible. Chapter 6 provides some illustrative applications to Programme "Education", using the example of the *Success for You* project. The examples focus on teacher training and general support for students, and show how oversubscription, phase-in and regression discontinuity designs could be used in this context. This section outlines some more general guidelines for selecting activities that can be evaluated via experimental or quasi-experimental methods.

Suitable types of activities

Experimental or quasi-experimental methods are most promising when:

- The activity is well-defined and focused. For example, a teacher-training programme, a new mentoring scheme, an information campaign, or provision of digital or physical resources.
- Exposure naturally varies at the level of individuals or small units (i.e. some students, classes, teachers, schools or small geographical units receive the activity, others do not).
- The activity can be rolled out gradually, or eligibility can be defined through clear rules, such as score cut-offs, age thresholds, or limited capacity, creating a natural comparison group.
- Primary outcomes can be measured using existing or planned data sources for both exposed and non-exposed groups, such as student assessment scores, attendance completion rates, employment outcomes).

These conditions are often met by activities such as targeted remedial teaching, counselling and guidance services, digital-skills training, or specific vocational education and training and higher education support schemes. European Commission guidance on counterfactual impact evaluation highlights these types of "discrete" interventions as good candidates for experimental or quasi-experimental designs (European Commission, 2021[2]).

By contrast, causal designs are typically not appropriate for:

- Large national reforms that change the rules or curriculum for all learners at once. This is principally because there is no clear comparison group.
- Complex procedures that cover different types of activities, which may have distinct content, approaches or delivery modes.

In these cases, descriptive monitoring, theory-based evaluation and qualitative enquiry are more valuable tools for understanding what is changing in practice. Counterfactual methods may still be applied to selected sub-components where the conditions outlined above are met.

Ethical and legal concerns

Concerns about ethical and legal constraints, regarding withholding or delaying support from parts of the eligible population, are common. In education, these questions can be particularly sensitive as they may involve delaying additional support for learners who already need help. In some cases, such concerns can be mitigated through careful and transparent design although they cannot always be fully resolved. One option is an "encouragement" design, where access is universal among eligible groups, but some individuals are more actively encouraged to participate than others, for example through more intensive information or reminders. Another option is an "oversubscription" design, where demand exceeds capacity and available places are allocated randomly among eligible applicants. In both cases, the evaluation uses the random or quasi-random variation in exposure to construct a counterfactual while the underlying eligibility rights remain unchanged (Gertler et al., 2016[3]; European Commission, 2021[2]).

5.3. Why are advance planning and data access essential?

The success of experimental and quasi-experimental evaluations is driven by factors typically determined long before data analysis starts. The key decisions – who is eligible, how places are allocated, what outcomes are measured, and how data can be accessed – need to be taken explicitly during programme design and early implementation (European Commission, 2019[4]; 2021[2]; Gertler et al., 2016[3]). As the Executive Agency is preparing the next programming period (2028–2034), this is a timely moment to factor such design choices into new procedures from the outset, where desired and appropriate. For activities

under the current programming period (2021-2027), any quasi-experimental work would likely need to operate within implementation rules and conditions that have already been set. Even in these cases, evaluation can still be strengthened by securing access to suitable outcome data for both participants and potential comparison groups.

If the Executive Agency wishes to use experimental or quasi-experimental methods for a particular activity, the following elements should be considered at the conceptual stage of the activity where possible:

- **Align the planned activity with a clear theory of change (ToC).** The ToC for the specific activity should spell out the key causal links and the main outcomes that matter (e.g. attendance, completion, achievement, employment). These become the focus of the causal evaluation.
- **Choose the most suitable methodology for a given activity design; this determines the comparison group.** The Executive Agency and the beneficiary should first identify which methodological design is most adapted to the given context. For example, a phased roll-out can be exploited in a randomised design, an eligibility score cut-off is suitable for a regression-discontinuity design, and a staggered introduction may lend itself to a difference-in-differences design. The chosen methodology largely determines the definition of the control group. The control group cannot be selected post-hoc for convenience reasons, as this generally does not provide a credible counterfactual. The European Commission provides worked examples specifically targeted to Managing Authorities of ESF+ programmes of how to apply different designs to a given context (European Commission, 2020[5]).
- **Ensure outcome data will be available for both groups.** Outcome indicators should be chosen so that data can be collected for both those who receive the intervention and those in the comparison group, ideally from existing administrative sources. As discussed in Chapter 4, this often requires access to individual-level data from registers, such as education information systems, social security, or employment registers, and the ability to link data sources for the same individuals in a secure way (European Commission, 2020[5]).
- **Secure legal and technical arrangements in advance.** Data-sharing agreements, unique identifiers for linkage, and clear time frames for data provision need to be in place before the intervention starts, especially when outcome data are needed several months or years later (e.g. employment six months after graduation). European Commission guidance for ESF+ Managing Authorities emphasises that counterfactual impact evaluations should be planned within the overall evaluation plan, and that design and data requirements must be anticipated early (European Commission, 2019[4]; 2021[2]).

5.4. Practical implications for the Executive Agency Programme “Education”

For the Executive Agency, the main implication of this chapter is not that all procedures should be evaluated with experimental or quasi-experimental methods. Rather, it is that:

- A small number of well-chosen activities under Programme “Education” could be candidates for such methods, especially where interventions are discrete, outcome data are available, and there is political and institutional support for rigorous testing (see Chapter 6 for some illustrative examples).
- To make such evaluations possible, design and data needs must be considered at the moment when activities are conceptualised, ToCs are drafted, and indicators and data access plans are defined. As the Executive Agency moves into preparations for the next programming period, this is a particularly good moment to factor these requirements into early programme design and to create favourable conditions for using experimental or quasi-experimental methods, where desired and appropriate.

- Strengthening data infrastructure and data access arrangements, as discussed in Chapter 4, will make it easier to use these methods in the future since most designs rely on high-quality administrative data for both treatment and comparison groups.

In summary, experimental and quasi-experimental methods should be seen as complementary tools within a broader evaluation toolbox. They can provide robust evidence on the causal effects of selected activities, but only when the necessary conditions are in place: clear ToCs, well-chosen outcome indicators, and secure access to outcome data for all relevant groups.

References

European Commission (2023), *Peer Review Experimental methods for impact evaluation of social inclusion policies: Summary Report*, DG Employment, Social Affairs and Inclusion, https://ec.europa.eu/social/BlobServlet?docId=27660&langId=en (accessed on 22 May 2026). [6]

European Commission (2021), *Design and Commissioning of Counterfactual Impact Evaluations – A Practical Guide for ESF Managing Authorities.*, https://op.europa.eu/en/publication-detail/-/publication/dd4a4fc7-42a3-11ec-89db-01aa75ed71a1 (accessed on 28 December 2025). [2]

European Commission (2020), *Counterfactual impact evaluation of European Social Fund interventions in practice - Guidance document for managing authorities*, Publications Office of the European Union, https://op.europa.eu/en/publication-detail/-/publication/f82c5fb8-bb40-11ea-811c-01aa75ed71a1/language-en (accessed on 22 May 2026). [5]

European Commission (2019), *Advanced counterfactual evaluation methods - Guidance document*, Publications Office of the European Union, https://data.europa.eu/doi/10.2767/464242 (accessed on 26 May 2025). [4]

Gertler, P. et al. (2016), *Impact evaluation in practice*, World Bank Pubications, https://publications.iadb.org/en/impact-evaluation-practice-second-edition (accessed on 22 May 2026). [3]

HM Treasury (2020), *Magenta Book: Central Government guidance on evaluation*, https://www.gov.uk/government/publications/the-magenta-book/magenta-book-central-government-guidance-on-evaluation-html#chapter-3-evaluation-methods (accessed on 28 December 2025). [1]

6 Applying principles of evaluation practice to Programme “Education”

This chapter applies the principles covered in earlier sections to selected procedures under Programme “Education” 2021-2027 to illustrate what they might look like when translated into concrete tools for programme design, monitoring and evaluation. The procedures were selected in consultation with the Executive Agency Programme “Education” (hereafter, Executive Agency) and reflect diverse challenges, objectives, and delivery modalities across the education sector. The purpose of this chapter is not to prescribe specific theories of change, indicators or evaluation designs for Bulgaria. Rather, it is to provide worked examples that help readers conceptualise how some of the recommendations in this report could be applied in practice.

The first section of this chapter applies the recommendations of Chapter 2 on theories of change (ToCs) to the procedure *Modernisation of Higher Education Institutions 2.0* (henceforth, *Modernisation of HEIs 2.0*). It presents two illustrative ToCs – one at procedure level and one at beneficiary level (in this example, a consortium of higher education institutions (HEIs) or work package).

The second section then shows how indicator frameworks can be derived from ToCs and provides an illustrative indicator framework for *Modernisation of HEIs 2.0*. The section demonstrates how indicators can follow logically from the ToC and illustrates practical design choices, such as using cohort-based reference groups rather than only participant-based indicators. It also provides examples of beneficiary-level indicators that can support simplified-cost financing and small-scale results-based financing. Building on these examples, the section concludes by offering brief feedback on the indicator set used in the monitoring questionnaire for the procedure *Supporting Centres of Vocational Excellence*.

The third section considers data access implications for the indicator types proposed, distinguishing between beneficiary-only monitoring and approaches that require system-wide data to enable broader progress tracking and comparisons with suitable reference groups.

The final section illustrates how selected experimental and quasi-experimental methods could be applied to specific activities within the *Success for You* project, focusing on implementation features that would make such approaches feasible.

6.1. Theory of change: Modernisation of Higher Education Institutions 2.0

The following section presents a reconstructed ToC for the *Modernisation of HEIs 2.0* procedure. Its purpose is to illustrate how the approach recommended in Chapter 2 can be applied in practice. Importantly, this ToC has been developed ex post, using the already defined set of eligible activities as a starting point and matching them to the main challenges in the higher education system. This is the inverse of the ideal sequence: to be most helpful, the conceptualisation of the ToC should proceed in parallel with, and inform, the design of the procedure itself, in particular the identification of challenges to be addressed, the choice and formulation of activities eligible for funding, and the selection of outcomes to be monitored.

Developing the ToC and the activity menu together is important for several reasons. It encourages programme teams to think carefully and from the outset about which challenges should be prioritised in each procedure or project, which types of activities are likely to address which challenges and through which mechanisms these activities are expected to work. It also requires explicit discussion of the conditions that need to be in place for these mechanisms to function as intended, such as staffing, timetabling, digital infrastructure, or employer engagement. Finally, it provides a structured basis for choosing a limited set of indicators that can track progress towards intermediate and longer-term outcomes, rather than focusing only on activity delivery.

The section distinguishes between two levels. First, it sets out a procedure-level ToC that shows how broad activity clusters under *Modernisation of HEIs 2.0* are expected to contribute to national higher education objectives. Second, it illustrates what a beneficiary-level ToC could look like for an individual project proposal, submitted by one lead HEI and its partner institutions, where specific activities are linked to concrete outputs, intermediate outcomes, assumptions and risks. It is at this beneficiary or proposal level that a ToC tends to have the strongest practical implications, because the links between specific activities and different types of effects can be examined in detail. At this level, project teams can consider not only the intended mechanisms through which activities may improve teaching and learning but also possible unintended effects that may need to be monitored and managed. Finally, this section aims to show how these different ToCs can be aligned with each other, and how they can support both strategic planning at programme level and more detailed reflection on mechanisms, conditions and monitoring at the level of individual projects.

Theory of change: Procedure level

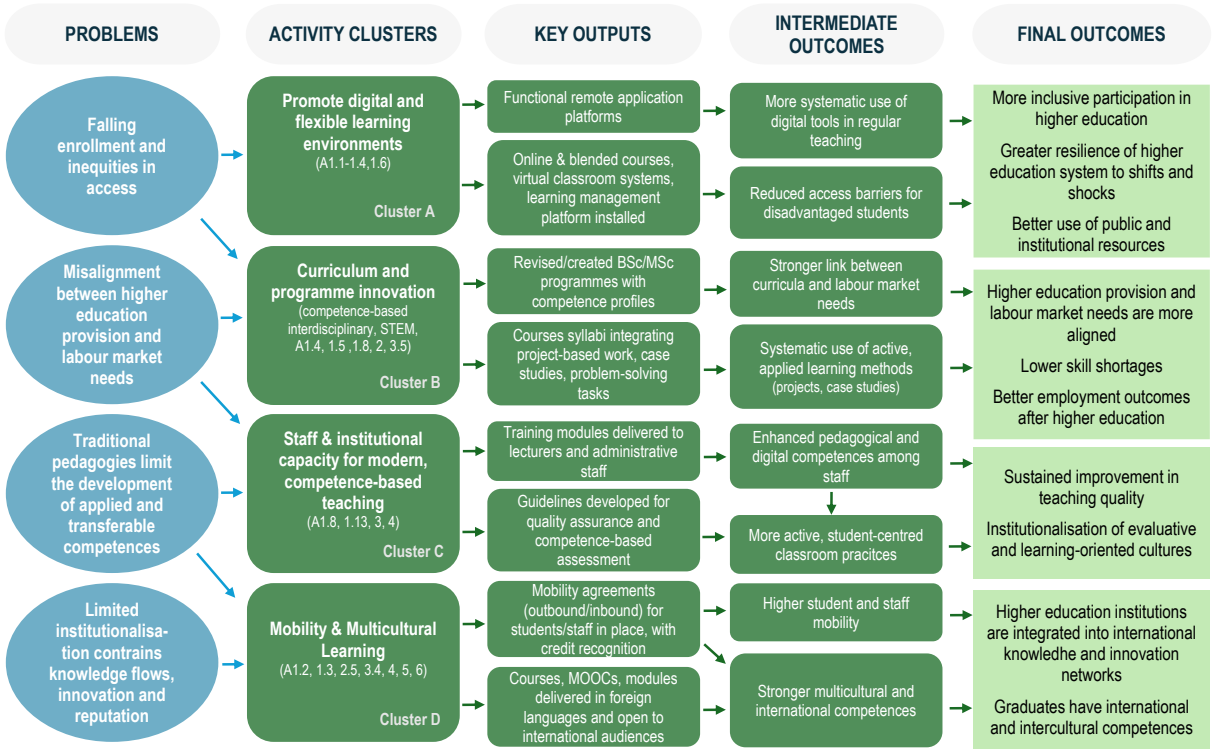
Figure 6.1 sets out a procedure-level theory of change for *Modernisation of HEIs 2.0*, structured around four broad clusters of eligible activities. For each cluster, it links the main problem area to the supported activities, expected outputs, intermediate outcomes and longer-term outcomes, and highlights a small number of practical assumptions that need to hold for the results chain to function as intended.

Activity Cluster A: Digital and flexible learning environments

Modernisation of HEIs 2.0 responds to the fact that participation in Bulgarian higher education remains uneven, and that traditional, lecture-based provision can be difficult to reconcile with work, caring responsibilities or residence far from a campus. The procedure therefore supports activities that develop remote application systems (Activity 1.1), digital and distance learning resources (Activity 1.2), Massive Open Online Courses (MOOCs) (Activity 1.3), virtual classrooms and digitally equipped science, technology, engineering, and mathematics (STEM) laboratories (Activities 1.4 and 1.6). These activities are expected to produce concrete outputs such as functioning online application platforms, sets of online and blended courses with interactive elements, and digital teaching spaces that are ready for regular use.

If these outputs are in place, and if teachers have time and support to adopt them, the intermediate outcomes are a more systematic use of digital tools in everyday teaching and reduced practical or logistical access barriers for some groups of students. In the longer term, this should contribute to more inclusive and efficient participation in higher education and to a system that can adapt its delivery modes more easily as needs change. Key assumptions are that students have access to suitable devices and stable internet connections, that digital platforms and equipment function reliably and are maintained, and that assessment practices keep cheating and plagiarism in check in online formats. If these conditions do not hold, digital provision may remain underused, mainly benefit more advantaged students or undermine trust in the quality of higher education.

Figure 6.1. Procedure-level theory of change for *Modernisation of Higher Education Institutions 2.0*



Key assumptions and risks (selection):

Operational conditions

- Digital platforms and equipment are installed, maintained and interoperable with existing systems.
- Classrooms, laboratories and ICT facilities are adequate for interactive and digital teaching.
- Teachers' workloads and timetables include protected time for course design, new assessment formats and participation in training.
- Participation in pedagogical and digital training is recorded and taken into account in staff appraisal.
- Academic calendars and timetables are adjusted so that project work, interdisciplinary modules and mobility fit without overloading students.
- Administrative staff have sufficient time and skills to support digital systems, revised curricula and mobility procedures.
- Examination and grading procedures include measures to keep cheating and plagiarism in check, including online.

Participant-side conditions (students and staff)

- Students have reliable access to suitable devices and stable internet connections.
- Students receive clear and timely information about new course formats, support services and mobility options.
- Students can cover any additional costs linked to mobility or new formats, with available financial support where needed/
- Staff are willing to engage with new teaching methods, digital tools and international collaboration.
- Language support is available where courses or mobility foreign language use.

External conditions and contextual risks

- Partner institutions abroad are able and willing to implement agreed mobility and joint teaching arrangements.
- Employment and external partners are available to provide input to curricula, guest teaching or supervision of practical projects.
- Credit recognition procedures for external and mobility courses operate in a predicatable and timely way.

Note: For a full description of activities (A1-A6) see the main text accompanying this figure.
Source: Authors based on Executive Agency "Programme Education" (2025[1]), *Application Conditions, BG05SFPR001-3.006 "Modernisation of Higher Education Institutions 2.0"* (accessed on 22 December 2025).

Activity Cluster B – Curriculum and programme innovation

Employers in Bulgaria report persistent difficulties recruiting workers with the competences they need, while many higher education graduates work in jobs that do not require tertiary-level skills. *Modernisation*

of HEIs 2.0 seeks to address this misalignment by funding the design and redesign of programmes and courses, alongside new resources that support applied learning.

The procedure covers several types of activities: developing STEM materials for student training (Activity 1.4); creating new courses and expanding Bachelor's and Master's programmes with explicit competence profiles and interdisciplinary or STEM content (Activity 1.5); providing additional training for students to develop cognitive, non-cognitive, digital, civic and social competences and project-based learning (Activity 2); and organising inter-university conferences and seminars on new teaching methods (Activity 1.8). These are supported by guidance on preparation of competence-based programmes and examination materials (Activity 3.5).

Together, these activities are expected to produce concrete outputs: revised Bachelor and Master programmes with explicit competence profiles, course syllabi that embed applied learning tasks such as projects and case studies, new STEM and interdisciplinary modules and institutional templates that guide competence-based programme design and assessment.

Provided that programme teams use these tools to adjust content and assessment, and that timetables and regulations are adapted so that curricula do not become overloaded, the intermediate outcomes are programmes that are clearly structured around competences, make systematic use of active learning, and are revised with reference to labour-market information and employer input.

Over time, these changes should contribute to a closer match between graduate competences and occupational needs, lower skill shortages in the labour market, and better employment outcomes for higher education graduates.

These effects depend on several important assumptions: that HEIs have enough discretion within existing workload and regulatory constraints to remove or consolidate outdated content when adding new elements, that employers and external partners are available to provide feedback or guest teaching where needed, and that students perceive the reformed programmes as attractive and manageable. If these assumptions are not met, programmes risk becoming formally more complex without substantive change in teaching and learning or may simply fail to attract sufficient enrolment.

Activity Cluster C – Staff and institutional capacity for modern, competence-based teaching

Many of the desired changes in curriculum and pedagogy depend on what academic and administrative staff are able and willing to do in practice. *Modernisation of HEIs 2.0* therefore invests in several complementary activities: structured training for lecturers on interdisciplinary and STEM teaching, higher-order thinking and problem solving, teaching technologies, multicultural and multilingual work and competence-based programme and examination design (Activity 3); training for administrative staff to work in multicultural and multilingual environments (Activity 4); opportunities to organise and participate in national and international conferences and seminars on new teaching methods (Activity 1.8); and the creation or upgrading of quality assessment and control systems in student training (Activity 1.13).

The immediate outputs of these activities are training sessions delivered, academic and administrative staff trained, internal guidelines and frameworks produced, and – where possible – new or strengthened communities of practice or working groups on teaching innovation established. If staff have protected time to participate, and if departments create space for experimentation and peer feedback, the intermediate outcomes should include improved pedagogical and digital competences among lecturers, observable shifts in classroom practice towards more interactive and student-centred methods, and stronger institutional capacity to design and run competence-based programmes.

In the longer term, these changes are expected to contribute to sustained improvements in teaching quality across institutions and to the development of a more evaluative and learning-oriented organisational culture.

This logic relies on several important assumptions: that participation in training and improvements in teaching is recognised in appraisal and promotion processes; that class sizes and physical spaces permit interactive methods; and that quality assurance functions focus on improvement rather than compliance only. If these conditions are not satisfied, staff may attend training without changing their practice, innovative methods may remain confined to a small group of enthusiasts, and quality assurance may fail to support the intended pedagogical shift.

Activity Cluster D – International mobility, multicultural learning and attractiveness

Bulgarian higher education remains relatively weakly integrated into international networks, with low levels of short-term mobility and limited inflows of foreign students and staff, despite significant outward degree mobility. *Modernisation of HEIs 2.0* seeks to strengthen international integration by supporting a set of mutually reinforcing activities: outgoing and incoming mobility of students and teachers with credit recognition (Activity 5); the development of distance learning resources and courses in foreign languages, including English-language MOOCs (Activities 1.2 and 1.3); additional training to help students develop skills for working in multicultural and multilingual environments (Activity 2.5); training for lecturers and administrative staff to work in multicultural and multilingual environments (Activities 3.4 and 4); and promotional activities, student-led information events and open days (Activity 6).

The activities are expected to generate concrete outputs, including mobility agreements with partner institutions, foreign-language online courses accessible to international audiences, students and staff with training in intercultural competences, and outreach events that raise the visibility and attractiveness of Bulgarian HEIs. Under the assumption that academic calendars and curricula are flexible enough to incorporate mobility without extending study duration, that credit recognition procedures operate predictably and without excessive administrative burden, and that financial support for mobility is sufficient for students from different socio-economic backgrounds, the intermediate outcomes are an increase in credit-bearing mobility and stronger multicultural competences among staff and students.

Over time, these changes are expected to contribute to deeper integration of Bulgarian higher education into international knowledge networks, enhanced institutional reputation and attractiveness, and graduates who are better prepared to work in internationally exposed sectors. However, if the aforementioned assumptions are not met, mobility and international activities may remain concentrated in a small number of programmes or among better-off students; uptake may be close to zero in heavily sequenced curricula such as medicine; and the potential of international exposure to feed back into domestic teaching and learning may remain largely unrealised.

Theory of change: Beneficiary (or project) level

The following illustration turns to the level of the individual beneficiary or project proposal and sets out a ToC for a digitalisation-focused work package. A work package is a set of activities or sub-activities for which a beneficiary applies for funding. For this procedure, the beneficiary is a consortium of HEIs. Digital transformation measures are currently a prominent feature of many education reforms and interventions in Bulgaria, including *Modernisation of HEIs 2.0*. At the same time, they provide a clear illustration of an intervention that can generate not only the intended benefits for access, flexibility and internationalisation, but also a range of unintended - and potentially negative - effects on inclusion, engagement and learning.

The purpose of this beneficiary-level ToC is to show how a HEI consortium can think through these different mechanisms in a structured and explicit way. By specifying how digitalisation activities are expected to influence engagement, access and performance, under which conditions and through which potential side effects, the ToC can help project teams identify what needs to be monitored and which forms of support may be required to ensure that digitalisation strengthens equity and learning outcomes as opposed to weakening them.

Work package: Digital and flexible learning

For this illustrative work package, a group of HEIs introduces flexible, digitally enhanced provision in several large Bachelor programmes, aimed at improving access, supporting learning and opening-up international opportunities. The work package is expected to contribute to more inclusive participation and completion in higher education, a more resilient and adaptable provision model, better use of digital investments, and stronger international integration and competences among graduates.

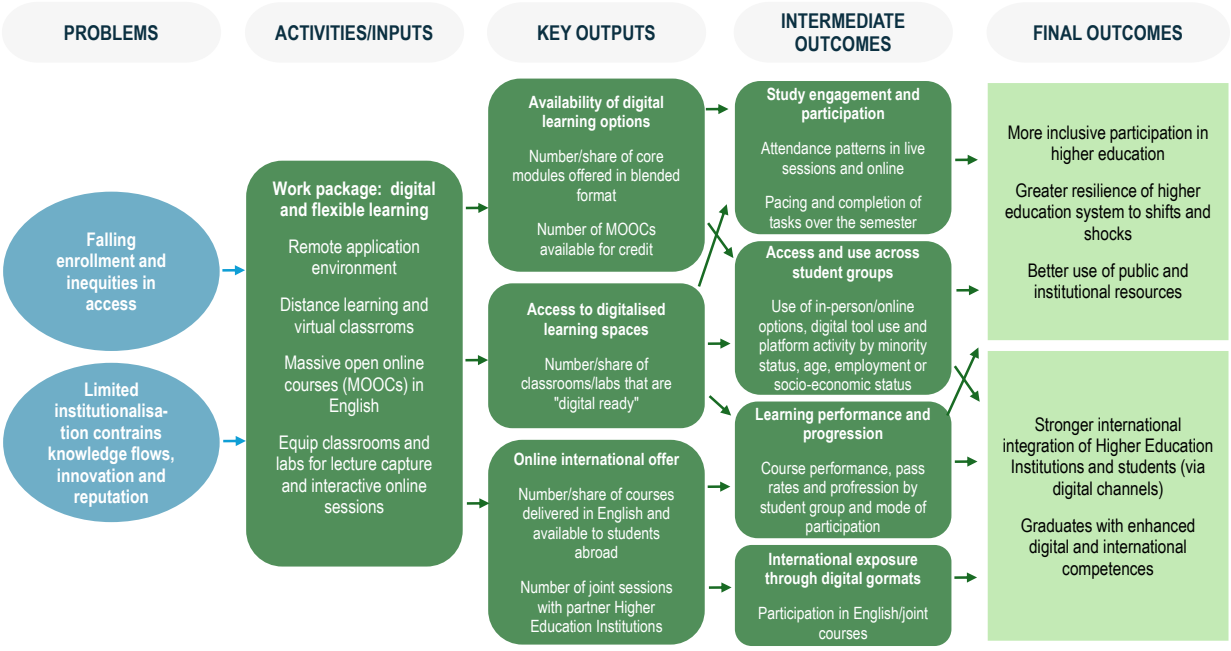
The activities focus on a coherent set of digitalisation measures. Institutions develop remote application and enrolment systems that allow students to apply and register without travelling to campus (Activity 1.1). They create distance and blended learning resources and virtual classrooms for core modules in targeted programmes (Activity 1.2) and produce a small number of MOOCs and open online courses in English that can also count for credit for enrolled students (Activity 1.3). They develop new digital learning resources for key STEM or quantitative courses where students often face difficulties (Activity 1.4) and equip selected classrooms and laboratories so that lectures and tutorials can be captured and delivered in interactive blended mode (Activity 1.6). Taken together, these activities generate three concrete outputs: 1) a portfolio of core modules available in blended or online format; 2) a catalogue of open online courses mapped to the internal credit structure; and 3) a set of “digital ready” teaching spaces that support high-quality audio-visual delivery.

How these investments translate into results for students and institutions depends on several intermediate outcomes, which can move in positive or negative directions. A first intermediate outcome are study engagement and participation patterns. In a favourable scenario, flexible digital offers make it easier for working, mature and geographically distant students to engage with their studies: travel and time costs fall, recorded lectures allow them to catch up when work schedules clash with live sessions, and online discussions provide an additional channel for questions. However, there are also several mechanisms that can weaken engagement. Online environments make self-regulation more demanding. The availability of recordings can tempt students to delay or skip live sessions, expecting to watch them later, and digital environments are typically full of competing distractions. If these mechanisms dominate, live attendance may fall, participation in interactive elements may weaken, and study time may become more irregular.

To understand which pattern prevails, institutions need to monitor attendance in live sessions, usage of recordings over time, and completion of online tasks, ideally by student group. Course design and institutional rules can then be adjusted, for example by setting clear pacing requirements, integrating graded online participation, and signalling where presence in interactive sessions remains important.

A second intermediate outcome is access and use across different student groups. Digitalisation can expand access for some groups. Remote application and online modules can lower barriers for students who live far from campus, have care responsibilities or combine study with employment. Students who are already digitally confident and have good equipment may use the new offers intensively. At the same time, digital access and skills are highly unequal. Students from low income or minority backgrounds may lack a quiet study space at home, share a single device among several family members, or rely on unstable internet connections. In the Bulgarian context, this can be particularly relevant for students from large households, where home environments can be noisy and crowded. Some groups, such as mature or first-generation students, may also have weaker digital literacy skills and feel less confident when navigating platforms or resolving technical issues. In such cases, a shift towards online provision can widen rather than narrow gaps in participation and achievement (Qaribilla, Indrajaya and Mayawati, 2024[2]). It is therefore important to track not only average use of digital offers, but also which student groups use them, disaggregated by socio-economic status, first-language background, age, employment and region. Where gaps are visible, institutions may need to provide device-loan schemes, dedicated on-campus learning spaces, and targeted digital-skills support to ensure that the work package does not unintentionally disadvantage vulnerable groups.

Figure 6.2. Beneficiary-level theory of change for the Modernisation of Higher Education Institutions 2.0 procedure



Key assumptions and risks (selection):

Operational conditions

- Digital platforms and equipment are installed, maintained and interoperable with existing systems.
- "Digital ready" classrooms and labs are available when needed and support good audio-visual quality.
- Teachers' workloads and timetables include time for redesigning courses and assessments for blended/online delivery.
- Examination and grading procedures include measures to keep cheating and plagiarism in check, including online.

Participant-side conditions (students and staff)

- Students have practical access to suitable devices, stable internet and a usable learning space at home or on campus.
- Students and especially mature or first-generation students have sufficient digital skills, or receive support to use the platforms.
- Students understand expectations for blended study and have support for planning and self-regulation.
- Staff are willing and able to use the digital tools provided, not only upload static materials.

External conditions and contextual risks

- National rules and institutional regulations fully recognise blended/ online modules and external MOOCs for credit.
- International partners are able to participate in joint online activities where planned.

Source: Authors based on Executive Agency "Programme Education" (2025[1]), *Application Conditions, BG05SFPR001-3.006 "Modernisation of Higher Education 2.0"* (accessed on 11 December 2025).

A third intermediate outcome is learning performance and progression in the digitalised modules. Digital resources can, in principle, support learning. Recordings and interactive materials can provide additional opportunities for spaced repetition and review, especially in demanding STEM and quantitative subjects. Online quizzes and assignments can offer frequent feedback and learning management systems can generate data for early-warning systems that identify students at risk of failing or dropping out. At the same time, evidence from other contexts suggests that fully online or poorly structured blended provision can be associated with lower course performance and higher dropout, especially for students with weaker prior preparation or lower self-regulation (Edwards and Clinton, 2018[3]). If lecture capture primarily substitutes for engagement in interactive sessions, or if students rely heavily on last-minute viewing of recordings, the potential gains from repetition may not materialise. To see which mechanisms are at work, institutions need to compare pass rates, grades and progression patterns in digitalised modules with those in face-to-face modules, and to examine differences by student background and mode of participation. This

can inform decisions about which types of courses are suitable for blended or online delivery, how to combine recordings with active learning elements, and which students need additional support with planning and study habits.

A fourth intermediate outcome is international exposure through digital formats. The introduction of English-language MOOCs and open online courses, together with joint virtual sessions with foreign partners, can provide students with more opportunities to interact with international peers and teachers. This can strengthen intercultural competences and familiarity with international standards in their field. However, international digital offers may mainly be used by a subset of students who already have strong language skills and high confidence while others feel excluded by language barriers or feel that these courses are not "for them". There is also a risk that international contacts remain occasional and one-off, thus having a limited effect on learning, motivation and employability. Monitoring participation in international online components by field, language level and student group, and integrating such components into assessed modules rather than treating them as optional add-ons, can help make sure that the international dimension contributes meaningfully to graduates' competences and does not become an additional source of inequality.

If the work package is designed and supported with these mechanisms in mind, the intermediate outcomes can contribute to the longer-term goals. More inclusive patterns of engagement, access and learning performance can translate into more inclusive participation and completion in the targeted programmes, particularly for under-represented groups, working students and those living far from campus. If used well, high-quality digital provision can increase the resilience and adaptability of HEIs, allowing them to respond to shocks and changing student needs without compromising quality. Intensive use of "digital-ready" spaces and shared online resources can lead to better use of public and institutional resources invested in digitalisation. Finally, where international online courses and joint sessions become a regular part of study programmes, the work package can support stronger international integration of HEIs and students, and contribute to graduates with enhanced digital, international and intercultural competences.

Throughout, several assumptions need to be recognised and monitored. On the operational side, digital platforms and equipment must be reliable and integrated, timetables must allow for blended formats without inflating workloads, and examination practices must be adapted to keep cheating in check. On the participant side, students need practical access to devices, connectivity and usable learning spaces, and they need support to build the digital and self-regulation skills required for blended study. Staff must be willing and able to integrate digital tools into their pedagogy rather than simply uploading static materials. On the external side, national and institutional regulations must recognise blended and external online modules for credit, and partner institutions must be able to join joint online activities where planned. Making these conditions explicit as assumptions, and linking them to specific monitoring indicators, helps ensure that the flagship work package delivers on its potential benefits while limiting unintended negative effects.

6.2. Choice of indicators

Indicators at the procedure level (Modernisation of HEIs 2.0)

For *Modernisation of HEIs 2.0*, indicators can now be derived directly from the ToC. As highlighted in Chapter 2, each element ("box") in the ToC should have at least one indicator with a defined data source and a clear measurement window. At the procedure level, these indicators allow progress to be tracked consistently across beneficiaries – from outputs (what is produced), through intermediate outcomes (what changes in practice or behaviour), to final outcomes (contributions to system-level goals). Data can be drawn from administrative sources and beneficiary-reported information.

Intermediate outcomes aim to capture the mechanisms through which change occurs. They are framed neutrally so they can register movements in either direction – positive or negative. Importantly, definitions

of intermediate and final outcome indicators should not be inherently tied to programme activities. This ensures that indicators can, in principle, be computed for both participating and non-participating HEIs. Indicators for participating HEIs can then be compared against pre-programme values, national averages, non-participant HEIs, or specifically chosen counterfactual groups.

Annex A.1 provides a detailed menu of possible procedure-level indicators that could be aggregated across beneficiaries or participating institutions.

Indicators at the beneficiary level

Annex A.2 provides an example menu of indicators for the digitalisation work package presented in Figure 6.2, which could be used by a HEI (or consortium) for monitoring and reporting purposes, in agreement with the Executive Agency. As discussed in Chapter 3, indicators should be closely aligned with the beneficiary-level ToC, and serve three linked purposes:

- support internal monitoring and learning, using target bands and triggers for extra support or corrective action;
- provide a clear basis for simplified-cost reimbursement (for example, for 90 % of funding);
- allow for a small, results-based financing component (for example, for 10% of funding).

All indicators are defined at programme or institution level, so that they are comparable over time and, if needed, with non-participating HEIs.

The described indicator set makes it possible to minimise financial risk for beneficiary HEIs because most of the funding can remain on a simplified-cost basis linked to verifiable outputs (e.g. having platforms in place and a minimum coverage of blended modules). At the same time, attaching a modest share of results-based financing to a few simple engagement and performance indicators helps align institutional incentives with the policy objectives. It rewards high-quality implementation, as well as the sustained effort required to ensure that digital provision works effectively for different student groups. Used in this way, indicators are not only a compliance tool, but also a basis for internal monitoring and learning, signalling where engagement is weaker than expected, where gaps between student groups are emerging, and where course design, teacher practices or student-support arrangements may need to be adjusted during implementation.

6.3. Data access concerns

As the accessibility of data sources for the different indicators outlined in the previous section varies widely, this section briefly considers questions of data access. The purpose is to outline practical considerations and identify areas where targeted improvements could generate high value. The discussion is intended as inspiration for possible high-value steps forward. It is not meant to be prescriptive.

The section distinguishes between two broad approaches to assessing progress and outcomes. One approach focuses on beneficiary institutions only, relying primarily on routine programme documentation and before-after comparisons. A second approach draws on indicators that can be constructed for wider reference groups, including non-participating institutions, which can support control group designs and wider system monitoring. These approaches serve different purposes and can be used in a complementary manner to strengthen both accountability and learning.

Beneficiary-based progress assessment using programme documentation

Where quantitative assessment of impact relies only on data for programme beneficiaries – without information on any comparison group outside this sample frame - causal claims about quantitative impacts

are naturally more constrained because it is difficult to observe what would have happened in the absence of the programme. However, beneficiary-only approaches can still provide useful evidence on implementation and on short-term changes, for example by comparing indicators before and after project implementation, as long as major concurrent shocks - such as economic recessions or pandemics - are considered when interpreting trends. Even in beneficiary-based assessments, it remains useful to define indicators for stable cohorts or reference populations (e.g. all students in a given programme and year of study) rather than only for participants in project activities(see Chapter 2) because participant groups are typically not defined in the year before implementation and therefore do not provide a meaningful baseline for before-after comparisons.

A key practical advantage is that a substantial subset of the indicators discussed in the previous section can be populated through existing reporting channels. Beneficiary technical and financial reports can be used to track whether the main intended deliverables of *Modernisation of HEIs 2.0* have been achieved and taken up within institutions. Examples include the establishment or upgrading of remote application and enrolment environments, the development of digital and distance learning resources, MOOCs or virtual classroom capacity, the creation or revision of competence-based and interdisciplinary curricula, the development of guidance and internal tools for quality assurance and competence-based assessment, and staff participation in pedagogical or digital training. Beneficiary reporting can also support simple utilisation and reach indicators where institutions can extract routine data from their internal systems, such as the share of programmes that offer blended or online core modules, the number of courses using digital platforms and basic counts of students and staff participating in mobility or internationalisation activities supported by the projects. Used in this way, reporting serves both compliance and learning functions by highlighting implementation progress and early signals of uptake, and by identifying areas where additional support or corrective action may be needed.

Moving towards system-wide monitoring and reference-group evaluation

For evaluation designs that use counterfactuals through control groups (see Chapter 5), it is generally necessary to obtain indicators for broader groups of HEIs, including non-participants. Similarly, monitoring progress in the wider system requires indicators that are not limited to beneficiaries. In these settings, beneficiary reporting and funding conditionality cannot be relied upon. Instead, indicators need to draw on system-wide administrative data, recurring sector instruments, or other arrangements that cover both participating and non-participating institutions. The following subsection therefore presents examples of data infrastructure and data-sharing arrangements that could support wider system measurement and, where relevant, future evaluation designs using comparison groups.

In practice, such arrangements can be implemented either through bilateral data-sharing agreements between competent institutions, or through a secure service hosted by the National Statistical Institute (NSI) (see Chapter 4). Under an NSI route, the Ministry of Education and Science and relevant agencies would submit datasets or lists for matching. The NSI would link these datasets in a secure environment and release only de-identified, suitably aggregated outputs (e.g. by HEI, field and cohort) intended for statistical analysis and publication. This model can reduce administrative burden on data holders and minimise the circulation of identifiable data. The examples below illustrate what such arrangements could look like and what types of indicators they could provide. The three examples serve different purposes and are presented independently. They are intended as illustrations of potential steps that could be explored, depending on feasibility and monitoring priorities.

Using the Bulgarian University Ranking System as a system-wide data source

The Bulgarian University Ranking System (BURS) already compiles a range of indicators across institutions and professional fields, including metrics derived from student and employer surveys. Given that the system is developed under contract with the Ministry of Education and Science, one option would

be to set up a clear arrangement which allows the Executive Agency to use a well-defined and stable subset of BURS outputs as system-wide measures, comparable over time and not limited to *Modernisation of HEIs 2.0* beneficiaries. Any access to underlying individual-level data could be defined on a clear public-interest legal basis under the European Union's General Data Protection Regulation and national law, with appropriate safeguards.

A related option is to explore whether BURS survey instruments could incorporate a simple, policy-relevant inclusion proxy that supports the types of disaggregated indicators proposed in Annex A. One practical example is language background, which could complement existing practice on "vulnerable and marginalised groups" while avoiding sensitive identity categories. Such a proxy would make it easier to monitor whether participation and outcomes in higher education evolve evenly across groups, and to interpret digitalisation and internationalisation measures through an inclusion lens.

Indicators that could be derived through such an arrangement include:

- Employer-reported recruitment difficulties, reported by professional field and province, as a signal of labour-market pressure that complements administrative vacancy data and helps identify fields with persistent shortages.
- Employer views on graduates' preparedness and skill gaps, reported by professional field and province, as an input into monitoring curriculum relevance and assessing whether programme activities are closing perceived skills mismatches.
- Inclusive participation indicators aligned with *Modernisation HEIs 2.0* goals, such as the share of enrolled students in each field who fall into defined inclusion-relevant categories used in the system, and changes in these shares over time and across participant HEIs and potential control groups.
- Where an inclusion proxy such as language background is available, disaggregated trends in enrolment, progression or completion indicators by student group, which can support analysis of whether digital and flexible provision is associated with widening or narrowing gaps.

Operationally, the main requirement is agreement on which BURS outputs can be shared, and at what periodicity and level of disaggregation. Where outputs are based on survey components that are not run annually, Bulgaria may wish to focus on indicators that are either stable enough to be updated less frequently or that can be complemented by administrative sources.

Linking student and graduate records to employment and earnings outcomes using administrative data

A second option is to link higher education student and graduate records to administrative employment and earnings data. This can be implemented through a bilateral agreement between the Ministry of Education and Science and the relevant data holders, or through the NSI secure-service route, where the NSI performs linkage and returns de-identified data, potentially aggregated by institution, cohort, and field of study.

This approach would allow Bulgaria to monitor labour-market alignment and graduate outcomes without introducing new graduate tracking surveys, and it would generate indicators for both beneficiaries and non-beneficiaries, facilitating a potential impact evaluation.

Indicators that could be derived through such a linkage include:

- Graduate employment rates 12-18 months after graduation, by HEI and field of study, with consistent year-on-year measurement.
- Graduate earnings distributions (e.g. median earnings) at 12-18 months and 3 years after graduation, by HEI and field of study, as a proxy for labour-market returns.

- The share of graduates working in sectors related to their field of study, where sector codes are available, as a first-order alignment measure.
- Where occupation codes are available, the share of graduates employed in occupations corresponding to their field of study.

Operationally, the key requirements are agreement on identifiers for linkage, specification of cohorts and variables, and a governance arrangement that ensures secure handling and disclosure control.

Linking accreditation and quality assurance records to programme registries

A third option is to make more systematic use of accreditation and quality assurance information. This would imply a data-sharing arrangement with the National Evaluation and Accreditation Agency to obtain structured extracts of accreditation outcomes and, where available, scoring elements by institution and programme or professional field, together with the timing of accreditation cycles. These extracts can then be linked to the national programme registry held through National Centre for Information and Documentation, using stable identifiers for different institutions, fields and programmes in higher education to create a consistent panel of quality-related indicators across the higher education system.

A more ambitious, but still feasible, extension is to move beyond basic accreditation outcomes. Bulgaria may consider either encouraging the National Evaluation and Accreditation Agency to gradually increase the use of structured, comparable scoring elements for teaching quality and related dimensions, or conducting an independent coding exercise that translates key elements of narrative self-evaluation and accreditation reports into a structured format. The latter can be implemented as a time-limited exercise, focused on a small set of dimensions that are clearly relevant for *Modernisation of HEIs 2.0* and designed to minimise burden on institutions by relying on documents that already exist.

Indicators that could be derived through such an arrangement include:

- System-wide time series of accreditation outcomes and scores by HEI and professional field as a contextual indicator for teaching quality and programme governance.
- Where structured scoring is available or introduced, a teaching quality index derived from accreditation components that are designed to be comparable across institutions and over time.
- Where relevant elements can be coded from narrative reports, structured indicators on selected programme features, such as the presence of competence-based assessment approaches, the use of digital and blended teaching modalities, or the provision of English-taught and internationally co-delivered programmes.
- Medium-term indicators on programme renewal and reform dynamics, such as the share of programmes that introduce major revisions between accreditation cycles, interpreted in the context of Programme “Education” programming periods.

Operationally, the main requirements are agreement on extraction formats and identifiers, a transparent coding framework if narrative reports are used and a governance arrangement that ensures that any structured indicators derived from narrative sources are interpretable and comparable. Given that accreditation cycles are not annual, these indicators are most useful as medium-term measures and as a complement to administrative and survey-based indicators rather than as high-frequency monitoring metrics.

Together, these examples show how Bulgaria could strengthen system-wide measurement while limiting reliance on beneficiary-only reporting and avoiding extensive new surveys. The NSI secure matching route described in Chapter 4 would provide a practical mechanism to link sensitive datasets while protecting confidentiality and can be applied to several of the data integration steps outlined above.

6.4. How experimental and quasi-experimental designs could be applied to Programme “Education”

Chapter 5 discussed how experimental and quasi-experimental methods can be used as one possible approach to identify the causal impact of specific and well-defined activities, without replacing a broader, theory-based evaluation for the overall programme. The present section illustrates how such designs could be applied in practice, using selected activities in Programme “Education” as examples. This is intended as a practical illustration of how these approaches could be implemented should there be interest in using them in the future, either for Programme “Education” or for future programming cycles.

This section shows how small design choices in implementation can make experimental methods feasible, without changing the content of the intervention itself. Such designs are most useful when there is an interest in obtaining a formal, quantitative estimate of the impact of an intervention, for example, in cost-effectiveness considerations. Experimental designs and qualitative methods can provide complementary insights, with qualitative approaches helping to explain how and why activities succeed or fall short and offering a deeper understanding of participants’ experiences and the influence of complex contextual factors.

The focus here is on the *Success for You* project. In particular, the illustrations relate to teacher training under Activity 1 and to the provision of general support classes to students under Activity 4. Activity 1 finances training for pedagogical specialists and other staff, with the aim of improving their capacity to provide general and additional support for students’ personal development. Teachers volunteer for these trainings. Their direct motivation is professional development, and an indirect motivation is that the training qualifies them to provide paid general and additional support classes under Activity 4. Schools typically contract independent training providers and purchase a given number of places for their staff. The examples below show two ways in which experimental designs could be embedded in this delivery model: an oversubscription design at teacher level and a phase-in design at school level. Which approach, if any, is suitable in a certain context would need to be decided with implementers, taking account of operational and political constraints.

Teacher-level randomised controlled trial with oversubscription design

The first example applies when there is excess demand for training places from teachers. Suppose that, due to funding constraints or based on the expected need for general support classes, a school has only 10 places available for a particular training. However, 18 eligible teachers would like to participate to obtain the qualification. In such a case, a random tie-break among interested and equally eligible teachers can be a fair and transparent allocation rule, and at the same time creates the conditions for a randomised controlled trial.

A possible protocol would be:

- The school defines eligibility criteria for the training, such as minimum tenure at the school, relevant subject, or role.
- All teachers who meet these criteria and wish to participate register as candidates.
- If the number of candidates exceeds the number of places, the school allocates the available places by random selection among the registered teachers.
- Teachers who must attend for operational reasons (e.g. the principal, or specialised support coordinator) receive a place outside the random assignment and are not included in the evaluation sample.

The teachers who are selected using the mechanism form the treatment group and the eligible teachers who did not receive a place form the control group. Within a single school, these groups are usually too

small to support reliable statistical inference. However, if the same protocol is applied in many schools, the resulting data can be pooled into a larger national sample. Because assignment within each school is determined by random allocation, treated and control teachers differ only by chance. In large samples – as implied by the law of large numbers – their average characteristics will be very similar, so that any systematic differences in outcomes can be attributed to the training with high statistical power. The impact of the training can then be estimated by comparing the two groups on key outcomes of interest. These may include assessed teaching competences, observed teaching practices in their regular lessons, and attitudes and self-efficacy in working with vulnerable students or in providing socio-emotional support.

This design has two main advantages. First, it is procedurally fair: among equally eligible applicants, each teacher has the same chance of receiving a place, and decisions are not driven by personal relationships or informal bargaining. Second, the method remains valid and unaffected regardless of whether 13, 17 or 30 teachers apply for 10 places. The internal validity of the comparison is guaranteed by the random allocation. What does change with numbers is statistical power: schools where many teachers apply for a limited number of places contribute more information than schools with only a few applicants. By applying the same protocol across many schools and pooling the resulting data into a national sample, the evaluation can achieve sufficient statistical power to detect meaningful effects of the training.

At the same time, the design has clear requirements: there needs to be a surplus of demand; schools need to adhere strictly to the protocol; only teachers who are allocated a place using this mechanism should attend the training within the study period; outcomes of interest need to be measured for all teachers in the sample, including those who did not receive a place, using administrative data and simple survey tools. It is also important to be explicit about what this design does and does not evaluate. It allows estimation of the effect of the teacher training on any outcome that can be observed for both treated and control teachers. This includes, for example, their assessed competences, classroom practices in regular lessons, attitudes and self-efficacy, and potentially the academic performance of students taught in their regular classes. However, it cannot be used to evaluate the impact of the general support classes themselves, because control teachers do not teach these classes. A separate design (see below) is needed to evaluate the effects of these classes on student outcomes.

School-level randomised controlled trial with phase-in design

In some contexts, there may be no excess demand at teacher level. Schools may be able and willing to support all teachers who are eligible and interested to be trained. Oversubscription designs are then not possible. In such cases, a randomly phased introduction of the training across schools can create a comparison group without denying access to any school in the long run.

A simple phase-in design would follow these steps:

- Define a group of eligible schools, for example all lower secondary schools in a certain region or all schools with a particular profile.
- Use random assignment to allocate each eligible school to one of three groups. The grouping should be truly random, using a clear and transparent procedure and not based on geography or any school characteristic (e.g. it is not sufficient to split into east, west and central Bulgaria, which is not a random sorting). With a sufficiently large number of schools, random assignment will spread different types of schools across the groups in a broadly even way - as implied by the law of large numbers. As a result, the groups become comparable on average, with a similar mix of urban and rural schools, different sizes and different student compositions.
- Training is then offered in phases by group. Group 1 receives the training in year 1, Group 2 in year 2, and Group 3 in year 3. Within each selected school, all teachers who meet the eligibility criteria and wish to participate can attend the training in the year assigned to their school.

- After year 1, compare outcomes in Group 1 (treated schools) with outcomes in Groups 2 and 3 (not yet treated schools). After year 2, compare Groups 1 and 2 (treated) with Group 3 (not yet treated). This is the core of the counterfactual evaluation: at each stage, the schools that have not yet been treated provide the counterfactual.

Here, the timing of access to the training is determined by random assignment at school level. This means that, before roll-out, the three groups of schools are similar on average in terms of their staff, students and environment. Differences in outcomes that appear in year 1 and year 2 can therefore be attributed to the earlier or later introduction of the training.

Because the design varies the training at the level of the whole school, it allows analysis of both teacher and student outcomes. For teachers, the same outcomes as above can be considered. For students, possible outcomes include test scores, grades, attendance, behavioural incidents, and indicators of risk of dropout. When comparing student outcomes, it is important to define student samples in a way that is consistent across treated and not-yet-treated schools.

The main advantage of the phase-in design is that no school is permanently excluded from the training. All eligible schools eventually receive access, and the evaluation uses differences in timing to create a valid comparison. This can be easier to justify than excluding some schools altogether. There are also clear limitations. Only short-term impacts can be observed. Once all schools have been trained (i.e. from year 3 onwards in the example), no untreated comparison schools remain, and long-term effects cannot be identified. The design also requires that implementation plans tolerate a staggered roll-out: in the example, only one third of eligible schools can start in year 1.

Similar designs may also be relevant in other parts of Programme “Education”. This includes interventions that are introduced to additional schools in successive school years, independent of the precise content of the intervention. One example is the procedure *Preventing harassment and violence and reducing school segregation* under the inclusive education theme of Programme “Education”. If the sequencing of this roll-out were designed as a random phase-in, the schools that have not yet entered the procedure would provide a clean counterfactual for those that started earlier, allowing inference on the causal impact of the activities. In contrast, where the timing of entry is determined based on school characteristics or administrative choices, such comparisons may be more difficult to interpret.

Taken together, the oversubscription and phase-in examples show that experimental evaluation is feasible for activities that have a similar structure to the *Success for You* teacher trainings, provided that allocation and timing rules are chosen with evaluation in mind and that, in the case of oversubscription, there is excess demand. They also underline that the content of the intervention - in this case, the teacher training - can remain unchanged. What matters for evaluation is who is eligible, how places are allocated when demand is high, and when different schools start. In the next subsection, a complementary example will consider how similar principles could be applied to evaluate the impact of general support activities for students under Activity 4 of *Success for You*.

The effect of providing general support: A regression discontinuity design

Activity 4 of *Success for You* finances general support for students’ personal development, often in the form of additional classes outside regular instruction hours. Eligibility for these support activities is determined by a diagnostic test administered by teachers, which assigns each student a score ranging from 0 to 18 points. Students who score below eight points on this diagnostic test – that is, students who do not reach a minimum proficiency threshold – are considered to be in need of support and therefore become eligible for general support classes. Students scoring eight points or above are not eligible. While there may be exceptions to this allocation rule in practice, this type of sharp, score-based eligibility rule creates a natural setting for a regression discontinuity design (see Chapter 5).

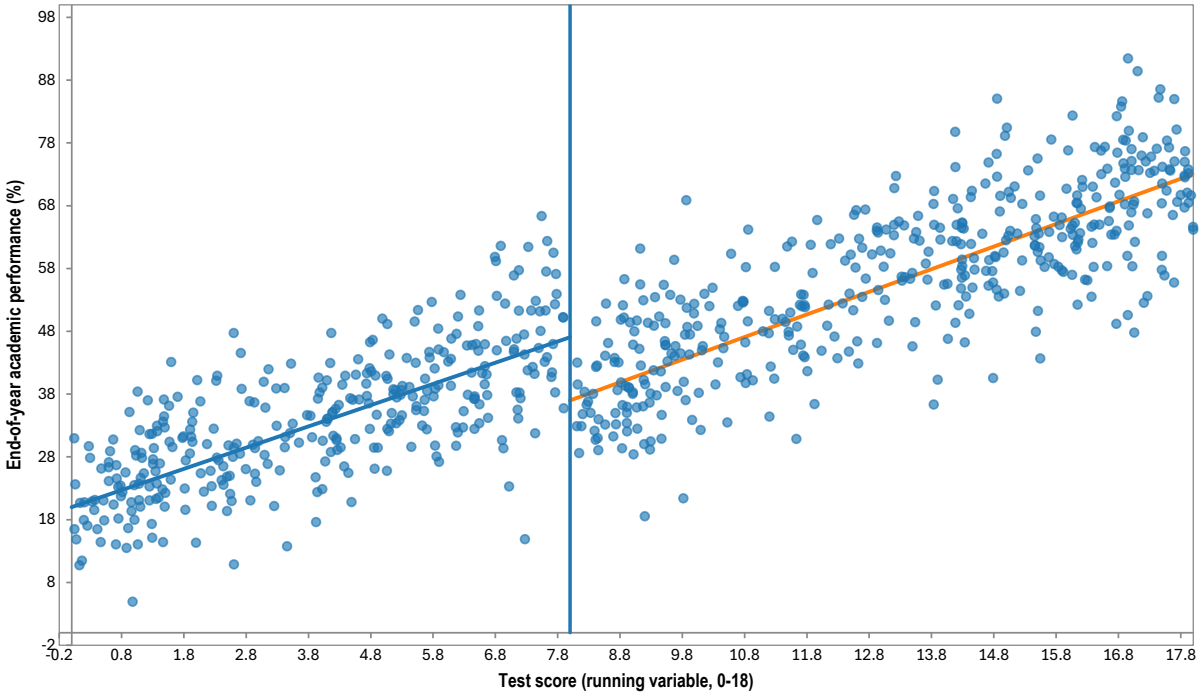
Put simply, the impact of receiving general support under Activity 4 can be estimated by comparing outcomes for students just below and just above the eligibility threshold. The key idea is that students who score seven and eight points on the diagnostic test are, in practice, likely to be very similar in underlying ability and background. The main systematic difference is that those with a score of seven qualify for general support, whereas those with a score of eight do not. By following both groups over time and comparing outcomes – such as academic performance at the end of the school year – any sharp difference in outcomes at the threshold can be attributed to the support rather than pre-existing differences. Figure 6.3 illustrates this logic using simulated data: each point represents one student, plotted according to their score on the initial diagnostic test (x-axis) and their academic performance at the end of the school year (y-axis, measured from 0 to 100). A positive impact of general support would appear as a jump at the cut-off, with end-of-year performance higher for students just below the score of eight than for students just above it.

Formally, regression discontinuity design estimates the relationship between the initial selection variable (here, the diagnostic test score) and a later outcome of interest (e.g. end-of-year test scores, grade progression, or an index of learning gaps). The size of the discontinuity in this relationship at the eligibility cut-off identifies the causal effect of the programme for students near that cut-off. Regression discontinuity design methods are powerful in this context because they can provide credible causal estimates without running an experiment: the diagnostic test and the cut-off rule are part of normal programme operation. The regression discontinuity design method relies crucially on a continuity assumption (see Chapter 5). While students below and above the threshold will differ in test scores and other characteristics, these differences should change gradually with the score. There is no substantive reason to expect a discrete jump in outcomes or student characteristics exactly at an arbitrary score of eight, in the absence of the programme. A jump that appears at this score, once the support rule is applied, can therefore be interpreted as the local impact of the support.

To make regression discontinuity design feasible for Activity 4 and similar future programmes, a few key features would need to be in place, or enforced where possible:

- **Clear eligibility rule:** Eligibility for general support should be defined by a transparent cut-off in a numeric score or index, such as a diagnostic test result below eight points.
- **Universal administration of the diagnostic test:** The diagnostic test should be administered to all potentially eligible students in the relevant grades, not only to students perceived as “at risk” of qualifying. This creates a well-defined reference group above and below the cut-off.
- **No other interventions at the same threshold:** The same cut-off should not trigger other changes in support or tracking, so that the only sharp change at the threshold is the probability of receiving general support.
- **Integrity and continuity of the diagnostic score:** There should be no systematic “manipulation” of scores (e.g. systematically rounding 7.5 up to 8). Simple checks of the score distribution can help detect and deter such behaviour.
- **Outcome data for all students in the reference group:** Outcomes of interest (e.g. end-of-year achievement, progression, attendance, early-warning indicators of dropout) should be available from administrative data for all students who took the diagnostic test, including those who are ineligible for general support and those who are eligible but do not actually attend support classes.

Figure 6.3. Evaluating general support under *Success for You*: A regression discontinuity design



Source: Author illustration using simulated data, based on Executive Agency "Programme Education" (2023[4]), *Application Conditions, BG05M2OP001-2.011-0001 "Success for You"*, https://iop.bg/s/Ukazaniya-Proekt-Uspeh-Za-Tebversiya-22025-1f.pdf (accessed on 29 December 2025).

When these conditions are met, a regression discontinuity design can estimate the impact of receiving general support for students at the margin of eligibility with strong internal validity, using data generated by the existing allocation rule. In the context of *Success for You*, this would provide evidence on whether the additional classes financed under Activity 4 help students close learning gaps and progress through school. It is important to note that this estimate is "local": it captures the effect for students near the cut-off and may not generalise to students with much lower or much higher initial scores. Nevertheless, in combination with other evidence and theory-based evaluation, regression discontinuity design offers a practical and robust way to assess whether general support activities are achieving their intended effect.

References

EAPE (2025), *Application conditions under the procedure for direct award of grant aid BG05SFPR001-3.006 "Modernisation of Higher Education 2.0"*, https://eumis2020.government.bg/bg/s/Procedure/Info/5cdb7749-ad19-4917-b58c-f9003525ed68 (accessed on 22 December 2025). [1]

EAPE (2023), *Application Conditions, BG05M2OP001-2.011-0001 "Success for You"*, https://iop.bg/s/Ukazaniya-Proekt-Uspeh-Za-Tebversiya-22025-1f.pdf (accessed on 29 December 2025). [4]

Edwards, M. and M. Clinton (2018), “A study exploring the impact of lecture capture availability and lecture capture usage on student attendance and attainment”, *Higher Education*, Vol. 77, pp. 403-421, https://doi.org/10.1007/s10734-018-0275-9. [3]

Qaribilla, R., K. Indrajaya and C. Mayawati (2024), “Digital Learning Inquality: The Role of Socioeconomic Status in Access to Online Education Resources”, *International Journal of Social and Human*, https://doi.org/10.59613/55gdmt96. [2]

Annex A. Indicators

A.1 Indicators at the procedure level

This annex provides a menu of possible procedure-level indicators for the theories of change outlined in Chapter 6.1. Indicators are grouped by the activity clusters defined in Chapter 6.1 and Figure 6.1. Indicators could usefully be aggregated across beneficiaries or participating institutions. A subset may be selected per activity cluster to keep the framework manageable.

Table 1. Output indicators (by activity cluster)

Activity cluster	Indicator	Definition	Data source
Activity Cluster A: Digital and flexible learning environments	Functional remote application systems	% of participating higher education institutions (HEIs) that operate a remote application and enrolment platform covering its core Bachelor and Master programmes	HEI technical reports, annual
	Share of programmes with digital options	% of modules offered in blended or online format in participating HEIs	HEI course catalogues or HEI technical reports/monitoring questionnaire, annual
	Availability of open online courses	Number and share of Massive Open Online Courses or open online courses developed under the procedure that are mapped to internal credit structures	HEI technical reports, annual
	Digital-ready teaching spaces	% of teaching rooms in participating HEIs that are equipped and certified as suitable for lecture capture and interactive online teaching	HEI infrastructure inventory, HEI technical reports, annual
Activity Cluster B: Curriculum and programme innovation	Competence-based programme design	% of Bachelor and Master programmes in participating HEIs that have documented competence-based learning outcomes and updated syllabi	Programme accreditation documents, syllabi, technical reports
	Course syllabi integrating project-based work, case studies, problem solving	% of courses whose syllabus includes at least one project-based assignment, case study or problem-solving task that contributes to formal assessment	Review of course syllabi submitted by HEIs
Activity Cluster C: Staff and institutional capacity	Staff participation in teaching innovation training	Average number of trainings in pedagogy, digital teaching or competence-based assessment attended, per staff member	HEI HR and training records, annual
	Existence of guidelines for quality assurance and competence-based assessment	% of HEIs that have approved written guidelines or frameworks for programme-level quality assurance and competence-based assessment	HEI documentation (senate or council decisions, quality assurance manuals)
Activity Cluster D: Internationalisation	Mobility arrangements in place	Number of active bilateral or multilateral agreements that generates student or staff mobility under the procedure.	HEI international offices, annual
	English-taught or internationally open courses	% of courses in participating HEIs that are taught in English or explicitly open to international students (including online).	Course catalogues, annual

Table 2. Intermediate outcome indicators

Area	Indicator	Definition	Data source
Study engagement and participation patterns (linked mainly to activity cluster A)	Patterns of attendance and online participation	% of students in a given programme who (a) attend at least X% of scheduled live sessions and (b) complete at least Y% of required online activities in core modules during the semester	Learning Management System (LMS) logs, attendance records, each semester
	Use of digital resources by student groups	Average number of lecture recordings viewed, and online learning objects accessed per enrolled student in core modules, with results disaggregated by programme and student characteristics (for example socio-economic status, minority status, age, employment status)	LMS logs linked to student registry, each semester.
	Completion and progression of under-represented groups	% of students who pass all core modules in the academic year, disaggregated by student group (socio-economic and minority status, age, employment status) and by whether modules use blended/online formats or not, compared to pre-programme cohorts	Higher education institution (HEI) academic records, annually.
Links between curricula and labour-market needs (linked mainly to activity cluster B)	Use of labour-market evidence in curriculum revision	% of programmes that report using at least one source of labour-market evidence (for example employer survey, sector council input, graduate tracer, Public Employment Service or Cedefop forecast) in designing or revising curricula in the last 2 years	Programme revision forms / quality assurance documentation, every 2 years.
	Students' perceived relevance of curriculum	% of students who report that the knowledge and skills from their programme are "highly relevant" or "rather relevant" to their intended career	Short student survey (for example integrated into standardised module evaluation or satisfaction surveys).
	Use of active and applied learning methods	Average % of the final grade which comes from project-based, case-based or problem-solving assignments, across compulsory courses in the current academic year	Course gradebooks or LMS assignment records, linked to syllabi.
Enhanced competences among staff and student-centred classroom practices (linked to activity cluster C)	Use of interactive digital tools by lecturers	Average number of different digital teaching tools or LMS features that teaching staff report using regularly (for example discussion forums, quizzes, polls, breakout rooms, collaborative documents, lecture capture)	Lecturer survey, every 1-2 years.
	Student perceptions of teaching quality in reformed modules	Average student ratings of clarity, perceived relevance, workload and support, by module	Student module evaluations, each semester.
	Student reports of active, student-centred teaching	% of students who report that, in a given module, they are asked to work in groups, discuss or solve problems, rather than only listen to lectures.	Student module evaluations, each semester.
	Observed active learning in teaching sessions (if feasible)	% of observed teaching sessions in which students engage in at least one activity beyond listening and note-taking (such as group work, discussion, problem solving, practical demonstration)	Structured classroom observations by trained peers, rolling sample each year (for example 10-20 % of staff observed).
International exposure through teaching and mobility (linked to activity cluster D)	Students' exposure to internationalised courses	% of students in a programme who complete at least one course with a substantial international component (for example English-medium, joint online sessions, virtual exchange) during their studies	Course transcripts and HEI reports, annually.
	Student participation in credit-bearing mobility	% of students in a given programme who complete a recognised credit-bearing mobility period (physical, blended or fully virtual) before graduation	Mobility office records linked to the student registry, for each cohort.
	Staff engagement in international teaching activities	% of academic staff who, in the current academic year, are involved in at least one international teaching activity (e.g. teaching on a joint course, co-developing a module with a foreign partner, delivering guest lectures to students abroad)	HEI international office and staff self-report survey.

Table 3. Final outcome indicators

Outcome area	Indicator	Definition	Data source
More inclusive participation in higher education	Inclusive enrolment	% of students from under-represented groups (low-income, minorities including Roma, rural, mature or working students) among all students enrolled in Bachelor and Master programmes in a given field or higher education institution (HEI)	National higher education statistics and HEI administrative data, annual
	Inclusive completion	Completion rate within the expected time for Bachelor and Master programmes, disaggregated by student group (for example low-income, Roma, rural, mature or working students)	National higher education statistics and HEI administrative data, annual
Greater resilience of the higher education system to shifts and shocks	Flexibility of delivery	% of credits in a given programme that can be delivered in blended or online format using institutional platforms, without changes to learning outcomes or accreditation status	Programme documentation and HEI self-report on delivery modes
Better use of public and institutional resources	Utilisation of digital teaching spaces	Average number of hours per week during teaching periods that "digital-ready" rooms and labs are used for teaching activities, divided by the total hours they are available	Room-booking systems and timetables
Higher education provision and labour-market needs are more aligned	Field-level alignment between graduates and employment	% of graduates in each field of study who are employed in occupations broadly corresponding to their field of study, 12-18 months after graduation	Graduate tracking survey and administrative employment data (for example social security records), with occupation codes
	Qualification-level alignment	% of employed graduates working in occupations that require higher education, 12-18 months after graduation	Graduate tracking survey and administrative employment data (for example social security records), with occupation codes
Lower skill shortages	Employer reports of recruitment difficulties	% of employers in who report "serious" difficulty in recruiting workers with the required skills at higher-education level, by field	Regular employer skills survey (for example national skills survey, Public Employment Service or chamber surveys)
	Hard-to-fill vacancies in higher education level occupations	% of vacancies in occupations typically requiring higher education that remain unfilled after a defined period (for example 60 or 90 days)	Public employment service vacancy statistics or labour-market observatory data
Better employment outcomes after higher education	Employment after graduation	% of graduates who are in paid employment or self-employment 12 months after graduation, by field and institution	Graduate tracking survey, social security or tax records; annual, by cohort
	Earnings after graduation	Median earnings of graduates 3 years after graduation, relative to the median earnings of workers with upper-secondary education of similar age	Graduate tracking survey, social security or tax records; annual, by cohort
Sustained improvement in teaching quality	Teaching quality in external reviews	% of programmes that receive a "good" or higher rating on teaching quality in external accreditation or evaluation exercises	Accreditation and external evaluation reports
Institutionalisation of evaluative and learning-oriented cultures	Regular internal evaluation of programmes	% of programmes that have undergone at least one structured internal evaluation based on data (for example student surveys, completion rates, graduate data) in the last 3 years	HEI quality assurance records
HEIs are integrated into international knowledge and innovation networks	Participation in international teaching and research networks	% of HEIs that are active members of at least one international network or partnership involving joint teaching, joint programmes, or joint research with a teaching component (for example European university alliances, thematic networks)	HEI international office reports and network membership lists
	Joint and co-taught programmes or courses	% of programmes that include at least one course jointly designed or co-taught with a foreign institution (including virtual or blended formats)	Programme documentation and international office records
Graduates have international and intercultural competences	Graduates' self-assessed international and intercultural competences	% of graduates who report medium or high levels of confidence in working in multicultural teams, communicating in at least one foreign language and collaborating across borders	Graduate survey with a short intercultural competence scale, 12-18 months after graduation

A.2 Indicators at the beneficiary level

The following section provides an example menu of indicators for the digitalisation work package presented in Chapter 6.2, which could be used by a higher education (HEI) or consortium for monitoring and reporting purposes, in agreement with the Executive Agency. As discussed in Chapter 3, indicators should be closely aligned with the beneficiary-level theory of change, and serve three linked purposes:

Support internal monitoring and learning, using target bands and triggers for extra support or corrective action;

Provide a clear basis for simplified-cost reimbursement (for example, for 90 % of funding);

Allow for a small, results-based financing component (for example, for 10% of funding).

All indicators are defined at programme or institution level, so that they are comparable over time and, if needed, with non-participating HEIs.

Table 4. Indicators framework for a work package on digital and flexible learning

Type	Indicator	Definition	Data source	Use
Output indicators (suitable for simplified-cost financing)	Functional remote application and enrolment system in place	Yes/No – the higher education institution (HEI) operates a remote application and enrolment platform covering all main Bachelor and Master programmes	HEI administrative report	Eligibility condition or trigger for releasing a first tranche of simplified-cost funding
	Share of programmes with digital options	% of modules offered in blended or online format	Course catalogue and Learning Management System (LMS) mapping, once per year	Main output indicator linked to simplified-cost financing (for example, 90 % of the work package budget once a minimum threshold is reached and verified)
	Use of "digital-ready" rooms for teaching	% of teaching hours in the timetable that take place in rooms certified as "digital ready" (equipped for lecture capture and interactive online sessions)	Room booking and timetable system, once per year	Checks that investments in equipment are actually used
Result-type indicators (suitable for internal monitoring, feedback loops, and small results-based financing components)	Student engagement in blended/online modules	% of students in programmes using blended or online formats who both (a) attend at least X % of scheduled live sessions (in person or live online) and (b) complete at least Y % of required online activities in core modules during the semester	LMS logs and attendance records, once per semester	Internal monitoring with target band (for example 70-85 %); triggers intervention or additional support if engagement falls below the band. Unlock a bonus payment (for example 10-20 % of the work package budget) if performance is above the band
	Equity in engagement with digital provision	Difference (in percentage points) between (i) the share of students from under-represented groups and (ii) the share of other students who meet the engagement criterion in indicator 4	Indicator 4 plus basic student background data (for example proxy for low income, minority status, age, employment), once per semester	Acts as an internal trigger for corrective action or targeted support measures if gaps exceed a defined band; bonus payments could be conditional on keeping the gap below a specified threshold to encourage inclusive implementation
	Pass rate in core digitalised modules	% of enrolled students who pass all core modules, in the current academic year	HEI academic records, annually	Supports internal monitoring of whether digitalisation contributes to or undermines academic performance; a bonus could be tied to maintaining or improving pass rates relative to baseline values

				While the primary interest is in modules offered in blended or online formats, conditioning incentives on pass rates only in digitalised modules may create adverse incentives for HEIs to exclude the most difficult modules from digitalisation. Such incentives can be kept in check by (i) pre-defining the list of core modules at the proposal stage, or (ii) using the change in pass rates from a pre-programme baseline within each module.