



The Urban Lab of
Europe!

PRACTICE

EDIT 04 MARCH 2021
BY UIA PERMANENT
SECRETARIAT &
ECORYS

Evaluation approaches

See on UIA
website



"Evaluation approach" is a conceptual analytical model, not a specific method or techniques that underlie and guide evaluation. What approaches are applied in

UIA projects?

The term 'evaluation approach' does not have a commonly agreed definition. For example, the Better Evaluation website – one of the most comprehensive evaluation resources available online – uses the term 'approaches' to refer to 'an integrated package of options', such as methods or processes. It illustrates this definition using an example of randomised controlled trials which 'use a combination of the options random sampling, control group and standardised indicators and measures'. We lean towards a broader formulation derived from the definition of a theory-based evaluation proposed by the [Treasury Board of Canada Secretariat](#):

Theory-based evaluation is an approach to evaluation (i.e., a conceptual analytical model) and not a specific method or technique. It is a way of structuring and undertaking analysis in an evaluation.

We would thus understand an 'evaluation approach' to be a conceptual analytical model, not a specific method or techniques. We see it as a way of structuring and undertaking analysis, but also data collection. It has to do with a set of principles and foci that underlie and guide the evaluation.

In the review of UIA M&E case studies, we asked ourselves a number of questions broadly related to evaluation approaches. Firstly, we wanted to know what approaches, including theory-based, were applied to evaluation and for what reasons. Secondly, we were interested in the link between the selected approach and the evaluation purpose and scope, as well as the overall design. While examining the evaluation processes, we looked closely at the applied evaluation criteria, questions and indicators, as well as interrelations between these elements. Last but not least, we analysed the involvement and roles of stakeholders in the applied approaches.

There are multiple evaluation approaches that one can follow. Sometimes approaches are distinguished by their purpose. These include outcome, impact or process evaluations. In terms of more conceptual approaches,

counterfactual approaches based on **experimental or quasi-experimental designs** are considered the most robust for determining whether an intervention worked.

Counterfactual

The counterfactual is a hypothetical situation describing "what would have happened had the project never taken place or what otherwise would have been true. For example, if a recent graduate of a labour training program becomes employed, is it a direct result of the program or would that individual have found work anyway? To determine the counterfactual, it is necessary to net out the effect of the interventions from other factors—a somewhat complex task."

Source: Baker, J.L., *Evaluating the Impact of Development Projects on Poverty. A Handbook for Practitioners*, The World Bank, 2001.

Experimental and quasi-experimental design

Experimental and quasi-experimental designs aim to test causal hypotheses, e.g. that a given change resulted from the project. Experiments differ from quasi-experimental designs in the way subjects are assigned to treatment. In true experiments, subjects are assigned to treatment randomly, e.g. by a roll of a dice or lottery, to eliminate selection bias. After the process of random selection, selected groups are then subjected to identical environmental conditions, while being exposed to different treatments (in treatment groups) or lack of treatment (in the control group). A quasi-experimental design lacks random assignment. Instead, assignment is done by means of self-selection (i.e. participants chose treatment themselves), administrator selection (e.g. by policymakers, teachers, etc.), or both.

To learn more about this approach, you can consult e.g.:

- White, H., Sabarwal, S., [Quasi-Experimental Design and Methods](#), UNICEF, 2014

They do not, however, answer questions such as why and how this was the case. When it comes to reaching conclusions about the 'why' and the 'how', **theory-driven approaches** are believed to offer better possibilities. They can take various forms, such as contribution analysis, utilisation-focused evaluation, realist evaluation, and more (see e.g. the concise descriptions of those theory-based approaches at [Better Evaluation](#)).

Theory-based approach

Theory-based evaluation has at its core two vital components. Conceptually, theory-based evaluations articulate a policy, programme or project theory, i.e. how activities are supposed to lead to results and impact, given specific assumptions and risks. Empirically, they seek to test this theory, to investigate whether, why or how interventions cause intended or observed results. Testing the theories can be done on the basis of existing or new data, both quantitative (experimental and non-experimental) and qualitative.

Source: European Commission, [Evalued Sourcebook – Method and techniques](#) 2013

For more information, visit e.g. [Better Evaluation website](#) or the website of the [Treasury Board of Canada Secretariat](#)

Some approaches, such as empowerment evaluation and participatory evaluation, are distinguished based on their emphasis on various interactions with stakeholders. Other approaches, such as **developmental evaluation**, prioritise learning, growing and developing knowledge over accountability functions.

Developmental evaluation

Developmental evaluation was pioneered by Michael Quinn Patton as a solution for innovative projects with multiple stakeholders, dynamic contexts and a level of uncertainty. It focuses on adaptive learning rather than external accountability, although the latter is still important. It aims to provide real-time feedback and generate learnings to inform development. The evaluator is an insider to the initiative and, apart from collecting and analysing data, can actively shape the course of the initiative providing information for decision-making and enabling learning.

Source: Dozois, E., Langlois, M., Blanchet-Cohen, N., [DE 201: A Practitioner's Guide to Developmental Evaluation](#)

To learn more about this approach, you can consult e.g.:

- Patton, M., Developmental evaluation applying complexity concepts to enhance innovation and use, New York, NY: Guilford Press, 2010.
- Patton, M., McKegg, K., Wehipeihana, N. (eds.), Developmental evaluation exemplars. Principles in practice, 2016.

Here, we discuss the lessons that we drew from evaluation approaches applied in the analysed UIA M&E case studies. We do not describe these approaches in-depth. So, for specific explanations of their defining features, we encourage you to visit websites such as [Better Evaluation](#), which can be a good starting point for further theoretical explorations in relevant literature.

Lesson #1: Have an evaluation approach

This lesson may be verging on a truism, but it is worth repeating nonetheless – have an evaluation approach. Just as with implementing any project or evaluation, the benefits of following a specific approach or combination are clear. It simply pays off to have a coherent vision of what the evaluation as a whole wants to and can deliver under specific project circumstances, and how this will be done in practice. Adopting an evaluation approach gives the evaluation a framework, a sense of purpose and coherence in collecting evidence and reaching useful findings that answer the right questions. After all, if you do not know where you are going, you may end up somewhere else.

A number of UIA M&E case studies applied theory-driven evaluation approaches based on the so-called ‘theory of change’ (e.g. [CURANT](#) and [U-RLP](#)). Other case studies also had theories of change (e.g. [Curing the Limbo](#) and [B-MINCOME](#)), but their role in evaluation was less visible. In Barcelona [B-MINCOME](#), we came across an example of a counterfactual evaluation based on an experimental design. However, the application of this approach was rather rare, especially in projects aimed at social change. In fact, as we explain later, some evaluators of the projects deliberately resigned from pursuing the counterfactual experimental approach either for political, ethical or practical reasons. There were also projects that followed a very flexible and open approach to evaluation, employing e.g. the action research paradigm (i.e. [CALICO](#) and [Curing the Limbo](#)) or additionally integrating specific research perspectives (e.g. **appreciative inquiry** in [CALICO](#)). Overall, very different levels of stakeholder participation could be observed. While none of the projects expressly adopted such approaches, it seems that at least some espoused the principles of developmental evaluation (e.g. [Curing the Limbo](#)). The latter was in fact intentionally done in Utrecht [U-RLP](#).

With the abundance of possible approaches to evaluation comes the difficulty of selecting the right one. To address this difficulty, while talking about [evaluation governance](#), we included two specific lessons – to engage a **strong research-focused evaluator** and to **involve the evaluator in the project** as early as possible. An experienced evaluator will facilitate the choice of an evaluation approach from the multitude of available options, making sure that it is sensitive to the project, both in terms of its ambition and working culture. They will also guide the team through the process of developing the approach and the overall evaluation design.

Appreciative inquiry

Appreciative inquiry is a research perspective that focuses on strengths, both of organisations and individuals. It intends to discover, understand and foster innovations in social-organisational arrangements and processes. As an approach, it is a response to negative deficit approaches, which primarily focus on defining the problems.

Source: Dawance, T., Smetcoren, A., Ryckewaert, M., Aernouts, N., De Donder, L., [Care and Living in Community, CALICO. Groundwork for evaluation and state-of-play](#), 2019, p. 57.

To learn more about this approach, you can consult e.g.:

- [Appreciative inquiry commons website](#)

Lesson #2

Align the evaluation approach with the project

“You need to make sure that your evaluation framework is coupled with your intervention logic. And the two of them are equally important.” (Source: BRIDGE representative)

The analysed UIA M&E case studies highlight the need to align the evaluation approach (and overall design) with the project. This lesson is grounded in recognition that evaluation is a judgement, a reflection on the project and its specific aspirations. So, what evaluation does or should do essentially depends on how project partners define activities, outputs, results and, most importantly, if one talks about a wider change, intended impact. Beyond that, the approach will also depend on what the implementing partners would like to learn about their project, e.g. whether they are interested in its effectiveness or rather the processual dimension, or indeed both.

“We didn’t want to answer questions that were not being asked in the first place.” (Source: B-MINCOME representative)

The approach also has to take into account the project’s specific contexts (social and/or political). Not all approaches will be possible or appropriate for all projects, as we explain in some of the lessons below.

For innovative projects, the alignment between the evaluation and evaluated intervention can – and possibly should – go even further. In the analysed UIA M&E case studies, evaluation functioned as an integral part of the projects with evaluators being involved in the activities early on (see Lesson #2 under [Evaluation governance](#)).

“I don’t remember a moment when evaluating [the project] was not on the table.” (Source: OASIS representative)

So, in the context of innovation, it is perhaps better to think of evaluation as inextricably linked to implementation, as underlined by representatives of Athens [Curing the Limbo](#). Frequently, and perhaps regrettably, in evaluation practice, this exercise is perceived rather as an addition to the project or something that has to be done for accountability, to present to a donor. As the analysed UIA M&E case studies show, this perspective cannot hold for innovative interventions which are, by definition, a test.

“I think that the most important issue about the B-MINCOME project is that really it was a project of evaluation of public policies. It was the aim of the project. It was not a project about innovation itself, but about evaluation of innovation taking several dimensions.” (Source: B-MINCOME representative)

The Barcelona [B-MINCOME](#) project is perhaps one of the most visible examples of how a project is a public policy intervention on the one hand and a social experiment on the other. This, of course, created a set of challenges for the partners and project management, so additional lessons can be drawn from those experiences. One of the lessons is that the boundaries of an experiment may be strongly affected by the legal framework for a given public policy. So researchers have to account for that limitation in their research design and analysis. Other UIA projects, such as Brussels [CALICO](#), were also implemented with the intention of gathering evidence for policymaking. And the material collected for the evaluations of Rotterdam [BRIDGE](#) and Utrecht [U-RLP](#) was used for follow-up or other public interventions.

Alignment of the evaluation with the project certainly benefits from evaluator’s early participation. The evaluator accompanies implementing partners in their co-creation processes and can make better sense of the project as a whole, while the partners feed into the evaluation approach and design (see Lesson #2 under [Evaluation governance](#)). This should translate into more fitting and robust approaches.

“[Being] involved since the very beginning helped me [...] to get better knowledge of the intervention activities and to progressively clarify what was considered as being a priority, as being particularly relevant from the point of view of the evaluation.” (Source: OASIS representative)

Ideally, the early stages of the project should proceed like a double helix with the project and evaluation developing in parallel yet connected streams.

“So shaping monitoring and shaping the content of the project activities kind of went hand in hand. [...] they [the external evaluator] were also reflecting with us what is realistic to implement, what is realistic to evaluate.” (Source: CoRE representative)

Lesson #3

Take time to develop the evaluation approach

It is no surprise that in complex and innovative projects, such as those supported by the UIA, there may be less clarity at the outset about their various components and aspirations. Consequently, enough time – possibly as much as half a year – should be allowed to develop underlying concepts, activities and the evaluation (its approach and the overall design).

In the Utrecht [U-RLP](#) project, which did not benefit from a preparatory stage added to the UIA project structure later on, the evaluators very strongly argued for its necessity (see Lesson #6 on theory-based approaches). Among the negative consequences, the lack of a preparatory stage delayed the evaluation's **baseline measurement**, since project activities started before the research tools were ready. This, in turn, could have compromised the end results, if not properly accounted for in the overall design.

“We’ve already said about the fact that the baseline was compromised so that was obviously a problem that we were constantly playing catch up from the word go, with everything actually.” (Source: U-RLP representative)

In some analysed UIA case studies, the initial development and design stage (for activities and evaluation) took as long as several months. One reason why a substantial amount of time may be necessary is that the concepts that the partners are initially grappling with can constitute the core of what the project is about. This was, for example, the case for Athens [Curing the Limbo](#), but in no way was this an exception. When looking back, representatives of the Athens project assessed that more time invested at the early stages, on creating solid grounds for common understanding of the project and evaluation, would have been beneficial.

“What has not been easy is that in the beginning we had this lofty definition that this is going to be an exchange programme, where the refugees will give back to the city. When we initially described this, it looked like a very good idea, but when we had to really pin down what will they give back, what are we giving them and what are we expecting from them, this became a big question that we grappled with along the implementation.” (Source: Curing the Limbo representative)

Moreover, UIA projects usually involve many partners which represent various sectors, including public authorities, academia, CSOs and businesses. With varying experiences, the parties can have different interpretations of the outcomes or a different understanding of the terms that underlie their cooperation. To bring the language, views and interpretations of many people to a common denominator is no small task. Since it is a prerequisite for a successful evaluation, it may become the evaluator's challenge to tackle. It can depend on the evaluator's position within the project and the selected approach, with theory-based approaches possibly demanding a lot in this respect. This kind of work was acknowledged as a challenge by some UIA evaluators, and not only in terms of time (e.g. [CURANT](#), [U-RLP](#)). In Antwerp [CURANT](#), the development of the project's 'change model' revealed differences in how partners viewed specific aspects of the project, yet its evaluator was not convince that its role was to bring the partners to a common denominator. Utrecht [U-RLP](#) evaluator put a lot of effort in building a common vision, while also acknowledging that some level of divergence will inevitably remain.

“Is this our role, to get everybody on the same page from the very beginning? We didn't see it necessarily like that.” (Source: CURANT representative)

Importantly, implementing partners are not the only ones who come to the project with a set of assumptions, expectations, interpretations or preferences. This is also the case for evaluators who share various backgrounds and may specialise in certain types of research, as was the case e.g. in Paris [OASIS](#). This can influence their leaning towards specific evaluation approaches and designs which may not be ideal from the perspective of implementing partners and the project. So, the time spent brainstorming and co-creating can also help evaluators to see beyond their own preferences and biases. It can help them to identify project partners' needs which the partners are not always able to articulate directly.

“[...] my main focus is on impact evaluation research, so on the overall outcome of the project. When I was contacted, I started to think more in terms of what are the outcomes [...] But, as soon as I was involved in the meetings [...], it was very clear that observing not only the final outcome, but the process was a priority, in terms also of learning lessons for other projects.” (Source: OASIS representative)

Even if all parties are on the same page, there are other issues important for the evaluation approach which may require time to determine before research begins. For once, the ideas that partners initially held about their projects and, consequently, evaluations may not stand the test of time. This was the case with some analysed UIA case studies related to asylum seeker reception and integration (i.e. [U-RLP](#), [CoRE](#), [CURANT](#)). So another important step in developing an evaluation approach concerns the verification of initial ideas against the possibilities on the ground. With often limited information at the application stage and a time lag between applying for and receiving funds, it is necessary to make time at the project's outset to do a 'reality check', so to speak.

“When you're designing these things they are on paper and [...] it's very difficult to see how it actually

works. [...] then the funding came through and we started talking through what was possible.”
(Source: U-RLP project representative)

The case of Rotterdam [BRIDGE](#) is an illustrative example of how the initial approach may need to be reconsidered once the funding comes through and activities begin. Few weeks after the project’s initiation, it became clear that the experimental approach selected at the proposal stage would not be feasible for political reasons (see Lesson #4). Importantly, the partners were able to elaborate a new satisfactory approach together with their evaluator.

“Whatever idea you have about the evaluation framework at the proposal stage will probably be amended during the execution of the project.” (Source: BRIDGE representative)

Finally, any initial concept notes, proposals or applications – whether they talk about project activities or evaluation – are necessarily limited in ‘detail’. So, it is obvious that these ‘details’ will have to be determined later on. We put ‘details’ in inverted commas, as we are not necessarily talking about small or inconsequential things. In research, small elements can potentially make a big difference.

“After the submission was sent and approved some of the nits and grits of the evaluation design were to be defined. So this was basically let’s say seven months of intense work and negotiations between partners to reach a system that would answer the questions that were of interest to the project.”
(Source: B-MINCOME project representative)

The preparatory stage gives partners and evaluators a chance to determine the right evaluation approach and design, including agreeing on evaluation questions, indicators and targets, data sources and data collection methods, and determining the availability of data, etc. It enables a plan to be developed and gives space for drafting documents which can guide the process, such as groundwork reports for Brussels [CALICO](#) and Antwerp [CURANT](#), a M&E plan for Rotterdam [BRIDGE](#) or Aveiro [Steam City](#), an evaluation manual for Athens [Curing the Limbo](#) or indicator matrices for Vienna [CoRE](#).

Baseline measurement

Baseline measurement describes the situation before a given intervention (e.g. a project or programme) begins. Baseline data shows values for indicators selected to measure performance, outcomes and impact of an intervention prior to its initiation. This data can be compared to data gathered throughout the intervention and after its completion to estimate change.

Source: Bamberger, M., [Reconstructing Baseline Data for Impact Evaluation and Results Measurement](#), The World Bank, 2010.

Lesson #4

Be aware of the advantages and limitations of the counterfactual approach

The counterfactual approach to evaluation aims to answer the question of whether an intervention worked, which makes it particularly suited for evaluating outcomes and impact. The key to this approach is a comparison between a treatment and a control group, which involves repeated measurements at least before and after the intervention’s implementation period. Thus, the approach can only be used if specific conditions are obtained, i.e. when there are people (or schools, households, districts, spaces) similar to the ones receiving the intervention, but that did not benefit from its services. The groups should be rather stable throughout the experiment and, preferably, should not experience distorting influences. These conditions can be hard to create. In exchange, the approach offers the most robust way of attributing causality between an intervention and the observed results, especially if it is based on experimental designs. The latter involves random assignment of people (or different units, e.g. households, schools, districts, spaces) to a treatment and control group.

The counterfactual approach based on an experimental design was used in the Barcelona [B-MINCOME](#) project. The intervention distributed different guaranteed minimum income (GMI) schemes to households in selected city neighbourhoods. Alongside GMI schemes, it also implemented a set of active policies. The counterfactual approach, not the only one applied in the project for evaluation purposes, was well-aligned with [B-MINCOME](#) specific circumstances, namely a predictable excess of demand for the project’s GMI offer. It was always clear to project partners that some people, even if they were eligible for support, would not be able to benefit from it due to a lack of sufficient resources within the project.

Implementation of the counterfactual approach in Barcelona [B-MINCOME](#) required very strong integration of the

public policy component with evaluation taking place from the beginning to create the right conditions for measurements to be possible. In fact, as project partners themselves note, [B-MINCOME](#) was both a different way of doing public policy and a social experiment at the same time. This, of course, was not without its implications and tensions were experienced between policy implementation and research.

“All the time, we faced the tension between the research and the evaluation and the structural policies and infrastructure, and human resources. [...] And sometimes this was not easy to manage.” (Source: B-MINCOME representative)

Barcelona [B-MINCOME](#) conducted three types of quantitative surveys for the purpose of its experiment in both treatment and control groups. It is a substantial data collection effort, generally characteristic of the approach, making it a costly endeavour. The need for counterfactual, extensive data collection and higher costs influence the transferability of the approach.

The counterfactual approach can be problematic for practical, ethical and political reasons, especially in interventions focused on social policies, involving people on the move or with dynamic life trajectories, e.g. asylum seekers. Some of the analysed UIA M&E case studies considered the counterfactual approach, but eventually decided against applying it in practice.

The partners in the Antwerp [CURANT](#) project decided not to apply the counterfactual experimental approach for practical reasons. The project catered to unaccompanied young adult refugees. During the design of the evaluation methodology, it was not clear whether the evaluation partner would have access to the group of young refugees who did not benefit from the intervention. In other words, the researchers did not know whether they would be able to follow the lives of these refugees to compare their situation with that of the refugees who received support from the project. Additionally, after the completion of recruitment interviews, the number of young refugees who were declared ineligible proved low.

Practical difficulties in establishing a control group coupled with ethical considerations led to the rejection of the counterfactual experimental approach by the evaluators in the Utrecht [U-RLP](#) project. The project targeted asylum seekers from the Overvecht asylum seeker centre and members of the neighbouring community. According to the [final evaluation report](#), the following challenges to the application of the counterfactual experimental design were identified:

- Participants experienced the project in different ways, to varying degrees and for different periods of time, which made comparisons difficult;
- Random assignment to experimental and control groups, i.e. provision of support to some asylum seekers and denial to others based on random selection, was highly problematic ethically in a project that targeted a vulnerable population;
- Asylum seekers themselves could choose to participate or not in the project and its parts.

Additionally, due to varying conditions in asylum seeker centres in the country, it was hard to find a centre that could function as a comparable control group for an experiment.

Finally, practical and political reasons prevented the application of the counterfactual experimental approach in Rotterdam [BRIDGE](#).

“In the design of the project, we wanted to do an almost proper scientific experiment [...] It only took us a few weeks before we learnt that this would be politically unacceptable.” (Source: BRIDGE representative)

The implementation of this approach in Rotterdam [BRIDGE](#) would mean that in the Rotterdam South district some schools would participate in the project and others would not. This was considered politically problematic in the school environment because some children in the same district would be denied support. Eventually, all students in all schools were able to take full advantage of the project. With the extensive use of **microdata**, the project evaluators were able to preserve some elements of the counterfactual. And the partners themselves were, in fact, more satisfied with the implemented approach than with their initial design.

“I think in our case, we were happier with the evaluation we did in practice than the evaluation approach we proposed in the proposal.” (Source: BRIDGE representative)

If you are interested in evaluating impact and would like to use the counterfactual approach, analyse all of the project’s circumstances early on to determine whether this is at all possible. Further down the line, consider whether it would be ethically and/or politically acceptable. If so, analyse whether you have sufficient resources to gather robust data. Finally, think about whether you are only interested in determining the presence (or lack) of

impact, or whether you want to understand why and how your intervention worked or did not work. The counterfactual evaluation will not provide many answers to the last two questions and will have to be complemented with other research efforts, as it was done in the Barcelona [B-MINCOME](#) project that we analysed.

Microdata

“Microdata are unit-level data obtained from sample surveys, censuses, and administrative systems. They provide information about characteristics of individual people or entities such as households, business enterprises, facilities, farms or even geographical areas such as villages or towns.”

Source: [World Bank](#)

Lesson #5

Consider a theory-based approach

If you are interested in your project's contribution to change, as well as answers to how and why your project works or does not work, consider choosing a **theory-based approach** to evaluation. This approach has been variously defined, but with a degree of simplification one can say that it is grounded in and follows an explicit project theory. The latter offers an explanation as to how activities are believed to lead to outputs, outputs to results and, eventually, how all these contribute to wider intended impact. Additionally, it provides information on the intervention's context, risks and assumptions. The project theory can also be called a theory of change or a change model. It can be presented in a number of visual forms. Below, we include some examples of project theories found in UIA M&E case studies. Importantly, theory-based approaches can accommodate various research designs, both qualitative and quantitative, including experimental.

A number of UIA M&E case studies, especially those focused on social policy, chose to implement an approach based on a theory of change (see also **Lesson #4** for a full background on how the approaches were selected):

- The evaluators for the Antwerp [CURANT](#) project opted for this approach, explaining the rationale for this choice in the '[Groundwork for evaluation and literature study](#)'. One of the important factors in the selection was recognising that this approach works for complex settings in innovative interventions which still require causal relations to be established.

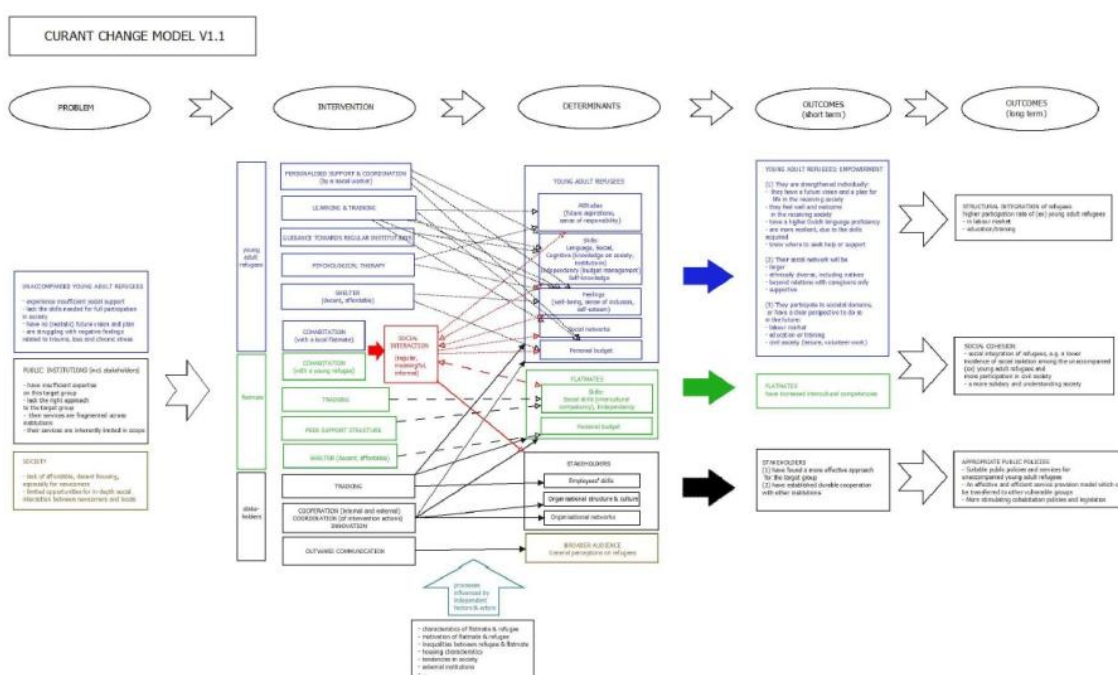


Figure 1: CURANT's change model. Source: CeMIS, [Groundwork for evaluation and literature study](#), May 2017.

- The evaluators of Rotterdam [BRIDGE](#) considered this approach suitable for their qualitative analysis and, as such,

complementary to the quantitative analysis of **microdata** that they carried out during the evaluation.

- The evaluators of Utrecht [U-RLP](#) decided to follow the theory of change approach because it was sensitive to the project's complex context and able to integrate the factors that actually made the experimental design impossible. The approach was flexible and collaborative enough to include various stakeholder perspectives and assumptions. In particular, it allowed for aligning and integrating partners' ideas about the project's objectives. It also offered enough space for learning to take place within the project. Ideally, however, due to [U-RLP](#) innovative nature, the evaluators would have followed a **developmental approach**, had it not been for the need to reconcile the emphasis on learning with the need for accountability. While the evaluation did not eventually follow the developmental approach, it did espouse its principles.

"[...] to my mind, this was a pilot project that was in development, that wasn't yet at the stage that clear objectives could be set out to the beginning and results indicators set, so that then we could measure and show progress in a linear fashion. So, I think in spirit, as evaluators [...] we had a very developmental evaluation approach as a philosophy [...] we wanted to have that supporters role as well [...]" (Source: U-RLP representative)

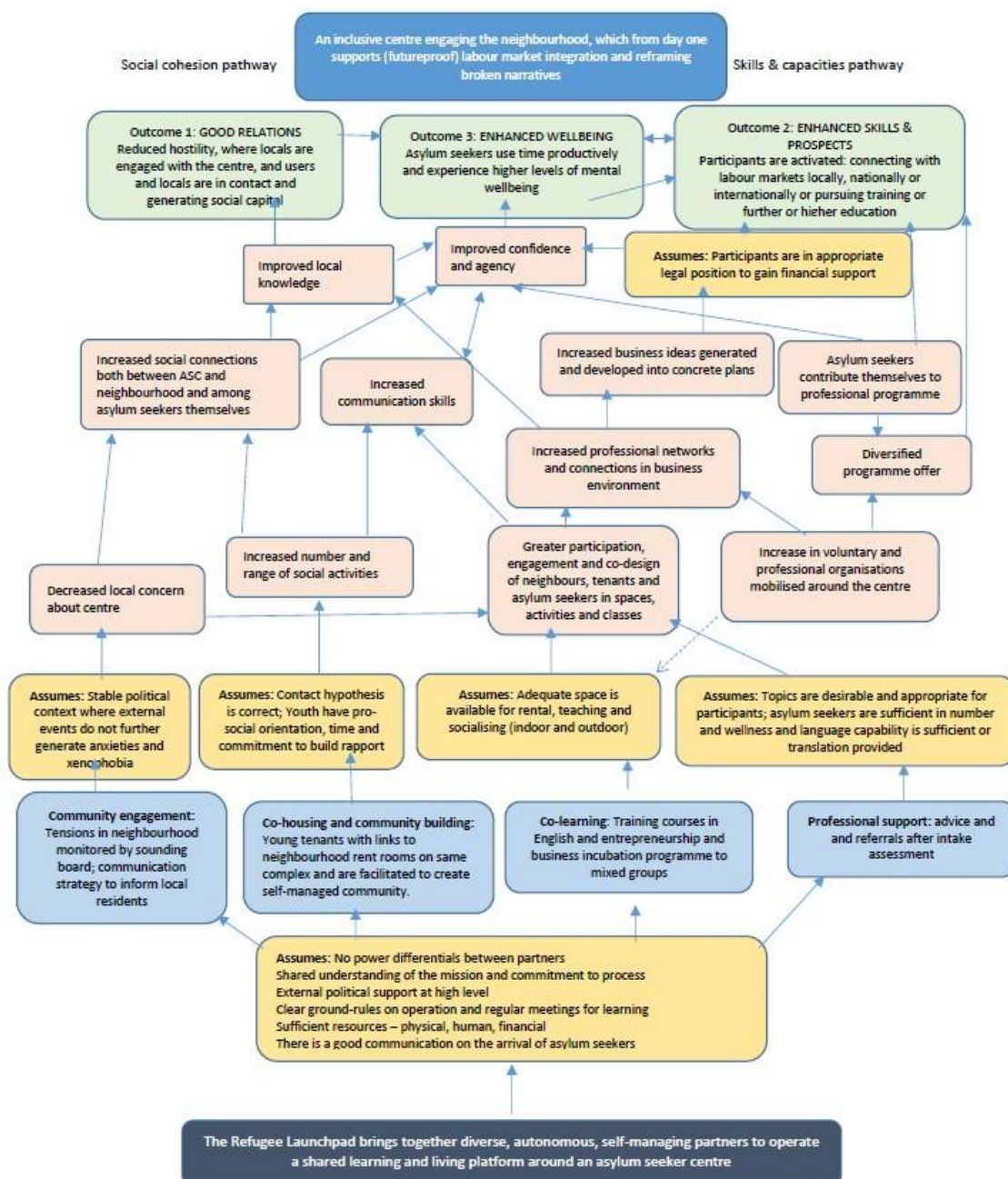


Figure 2: U-RLP's theory of change - end of project (October 2019). Source: Oliver, C., Dekker, R., & Geuijen, K. (2019). The Utrecht Refugee Launchpad: Final Evaluation Report. Oxford: COMPAS, University of Oxford.

Other projects also developed theories of change, but their use for evaluation was not clear at the stage of Ecorys' review (e.g. [Curing the Limbo](#), [CoRE](#) and [B-MINCOME](#)).

As evidenced by case studies such as Utrecht [U-RLP](#), developing a theory of change creates possibilities of reaching greater coherence among partners as to what the project is trying to achieve and how, as well as clarity around the external elements which can influence its effectiveness. Once developed, such theories can be a

versatile tool supporting the establishment of a monitoring system in the project. Later on, ideally, a theory should guide the evaluation process by influencing elements such as:

- the formulation of evaluation questions, so that they can test various elements of the theory;
- the design and implementation of data collection and analysis;
- the analysis and interpretation of evaluation results.

At the same time, one of the lessons (e.g. from the experiences of [U-RLP](#)) is that theory development can be very time-consuming. This is influenced by the multiple partner set up of UIA projects and their innovative and pilot character. The same reason as to why theory development is useful, e.g. lesser clarity among partners about possible project outcomes, also intensifies the challenge involved in theory creation. For this reason, project evaluators underlined the importance of allowing time for the preparatory stage of the evaluation process (see Lesson #3). This is particularly important if a theory of change is to be consistently used for monitoring the project, as well as designing and implementing the whole evaluation.

“If you want to do an evaluation of whether the aims have been achieved, you’ve got to have a clear understanding of what the aims are. And starting out with, I think, 57 aims – that’s problematic. [...] And my recollection is just how much time you’ve had to spend working through with everyone.”
(Source: U-RLP project representative)

Despite its flexibility, a theory of change approach can still be somewhat limiting for innovative projects, as observed by the Utrecht [U-RLP](#) evaluator and evidence by the rejections of this approach in Vienna [CoRE](#). Since the interventions have a pilot character, develop in dynamic circumstances involving high unpredictability, with many factors being outside partners’ control, it is difficult to establish at the outset a linear path of how things would work throughout the project. In this sense, a theory of change could be trying to prematurely solidify something that is essentially still fluid, and possibly better left that way for the time being.

“We would prefer a form of evaluation that is as much about developing the project and helping to identify those outcomes as you’re going through. Whereas the theory of change to some extent is trying to set that from the beginning.” (Source: U-RLP representative)

The evaluators from Antwerp [CURANT](#) pointed to another challenge involved in the theory of change approach related to how the theories are developed and, therefore, whose perspectives they represent. Theories of change capture stakeholders’ assumptions about actions required to solve a specific problem and reasons why these actions constitute a solution. In a nutshell, they outline how the stakeholders expect the projects to work. So, even if developed cooperatively with project partners, they can potentially re-create and perpetuate unjust institutionalised assumptions and biases. There is a danger that they will represent a top-down approach not rooted enough in the experiences and aspirations of project beneficiaries. For this reason, as underlined by Antwerp [CURANT](#) evaluators, the evaluators need to preserve a critical stance throughout the theory development process, focusing on the capabilities and agency of the beneficiaries. Re-examining stakeholder-led theories in light of up-to-date research can also be helpful in questioning some of the assumptions and biases.

...if refugees have access to an **individualized and integrated menu of services** and **interact with local active citizens groups** ...a sense of **belonging/openness and acceptance** is improved among both groups ...

THEORY OF CHANGE

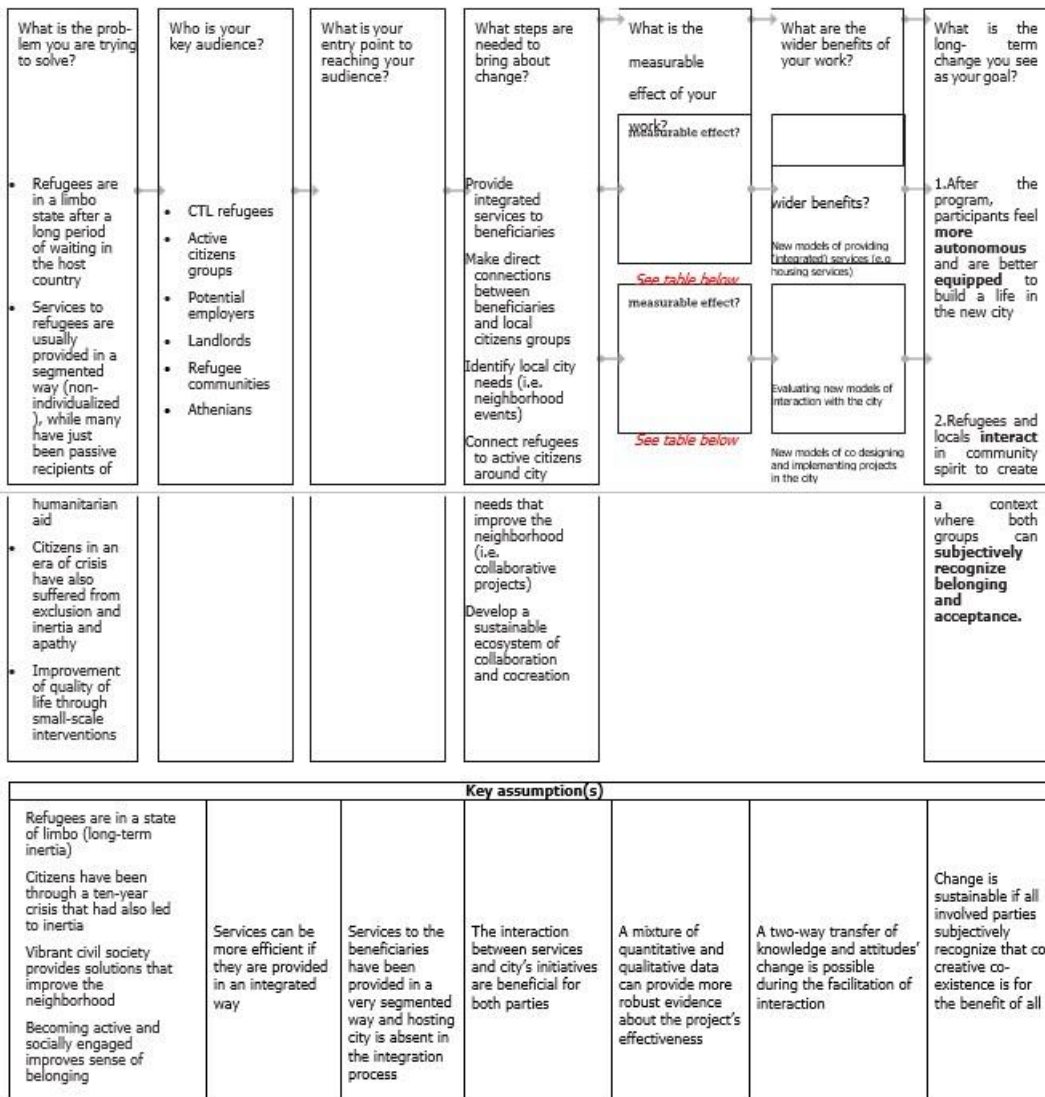


Figure 3: Athens Curing the Limbo's theory of change. Source: Curing the Limbo - Evaluation Handbook (V.3.1), February 2019.

Theory based approach

For more information, visit e.g. [Better Evaluation website](#) or the website of the [Treasury Board of Canada Secretariat](#).

Microdata

"Microdata are unit-level data obtained from sample surveys, censuses, and administrative systems. They provide information about characteristics of individual people or entities such as households, business enterprises, facilities, farms or even geographical areas such as villages or towns."

Source: [World Bank](#)

Developmental evaluation

Developmental evaluation was pioneered by Michael Quinn Patton as a solution for innovative projects with multiple stakeholders, dynamic contexts and a level of uncertainty. It focuses on adaptive learning rather than external accountability, although the latter is still important. It aims to provide real-time feedback and generate learnings to inform development. The evaluator is an insider to the initiative and, apart from collecting and

analysing data, can actively shape the course of the initiative providing information for decision-making and enabling learning.

Source: Dozois, E., Langlois, M., Blanchet-Cohen, N., [DE 201: A Practitioner's Guide to Developmental Evaluation](#) 2010.

To learn more about this approach, you can consult e.g.:

- Patton, M., Developmental evaluation applying complexity concepts to enhance innovation and use, New York, NY: Guilford Press, 2010.
- Patton, M., McKegg, K., Wehipeihana, N. (eds.), Developmental evaluation exemplars. Principles in practice, 2016.

Lesson #6

Consider action research as a possible research paradigm

Two UIA M&E case studies – Athens [Curing the Limbo](#) and Brussels [CALICO](#) – deliberately chose to conduct their evaluations within the **action research paradigm**. As defined in the evaluation plan for [Curing the Limbo](#), action research means ‘research informed by social action and leading to social action. Action is taken to improve practice and the research generates new knowledge about how and why the improvements came about.’ Thus, action research involves continuous weaving of knowledge creation with its use, observing the course of action and acting upon the observations. As the two UIA M&E case studies suggest, it is a very hands-on and attentive research practice. Implementation of this paradigm embeds learning/feedback loops in the evaluation structure. For this reason, the paradigm can be well-suited to evaluation of innovative projects whose trajectories and outcomes are not fully clarified, where not only accountability but also learning are key.

“[The evaluator] is giving as well. We’re presenting things. We are sharing. If there’s a question, we’re not just observing and looking at who is giving answers, then looking at the dynamics, but also participating.” (Source: CALICO representative)

Additionally, action research promotes stakeholder participation. At the practical level, it means that all partners are involved in the evaluation, as indeed was the case with both Athens [Curing the Limbo](#) and Brussels [CALICO](#). In Brussels [CALICO](#), the implementation of this paradigm benefited from the beginning from a strong leadership of its evaluators, i.e. researchers from the Vrije Universiteit Brussel. They developed a coherent evaluation framework (including the approach, research questions, data collection and analysis methods, stakeholder engagement, etc.), described in the [groundwork document](#). Similar leadership was initially missing in Athens [Curing the Limbo](#), but was later assumed by the project’s quality assurance team and considered vital to the evaluation exercise.

Being participatory in nature, action research also aspires to go beyond involving just the partners to account for inputs and feedback from other stakeholders, in particular beneficiaries. Ideally, stakeholder participation should also involve contributing to decision-making about the research agenda and the project’s course. In Brussels [CALICO](#), the researchers were aware of the dangers of ‘tokenist participation’, i.e. including the perspectives of the ‘usual suspects’. This is one of the traps that evaluators pursuing true participatory approaches need to take into account.

Importantly, the nature of action research endows the action researcher with significant power to take part in the project and influence its course. The evaluators of Brussels [CALICO](#) noted that the increased power that the researcher gains within this paradigm provoked ethical questions. With the ability to influence the course of action, also comes considerations of responsibility, weighing of interests and values.

“If we are in action research, it means we also have the power to steer things then. [...] So reflecting on the power we have as a researcher in action, research for me is an important ethical reflection.”
(Source: CALICO representative)

The power to stimulate action further highlights the need for appropriate subject-matter expertise, research skills and other competences on the part of an action researcher. Since action researchers are essentially insiders to the project, they should be able to rely on a body of knowledge to preserve a sufficiently critical stance throughout the project. This does not mean being critical towards the project, but rather being able to see the project in perspective. The nature of action research as a practice-oriented endeavour necessitates collection of robust evidentiary support. This is possible if the action researcher has strong research experience and sufficient soft skills to design and carry out data collection in cooperation with various stakeholders. The ability to build trust can at various points become crucial in action research, while trying to collect information, but also when presenting recommendations for actions.

Action research paradigm

To learn more about this approach, you can consult e.g.:

- Bradbury, H., The SAGE Handbook of Action Research, SAGE Publications Ltd., 2015.
- Coghlan, D., Doing Action Research in Your Own Organization, SAGE Publications Ltd., 2014, and the accompanying [website with tips and resources](#).

Lesson #7

Choose indicators wisely, aiming for quality not quantity

The issue of indicators captures, as if under a magnifying glass, some general tensions observed within the evaluations of projects such as UIA's. These are the tensions between the emphasis on measurement and accountability, on the one hand, and understanding and learning from pioneering interventions, on the other; tensions between quantitative and qualitative types of measurement and analysis.

The general lesson that we draw from these experiences is to strive for a balanced and pragmatic approach to indicators. At the most basic level, it means that the number of indicators should be reasonable. It is, however, hard to draw definite conclusions as to the exact numbers the cities should foresee, especially when interventions significantly differ in complexity. For example, Aveiro [Steam City](#) – which started out with several dozen indicators – managed to narrow those down to 26, yet the evaluators still considered this to be too many.

While the indicators should be treated with a degree of flexibility throughout a project, they should also be S.M.A.R.T., i.e. specific, measurable, achievable, relevant and time-bound. Though, importantly, focus should be given not only to their quantifiable character, but also to other characteristics of the S.M.A.R.T. approach. In particular, they should be specific to the intended results and outcomes to be able to realistically shed some light on their achievements, either in the shorter or longer-term perspective. This may be challenging at first, if the very outcomes pursued lack sufficient clarity. It is yet another reason to include a preparatory stage in the evaluation timeline (see Lesson #3 under Evaluation approaches). And, just like the understanding of outcomes, the indicators should be developed collaboratively.

“The meetings with the project partners were not just to validate indicators proposed by CEDES, but mainly to construct relevant indicators together. So every partner suggested some different indicators.” (Source: STEAM City representative)

The evaluators can play an important role in the process by translating the defined results into the right indicators and, subsequently, communicating and explaining those to the partners.

“We tried to translate these objectives, these goals into measurable outcomes and made clear [...] on the basis of which indicators the performance of the project would be evaluated or not. So the choice of the indicators was part of this collaborative process that started since the very beginning.” (Source: OASIS representative)

There is a benefit to be drawn from working on indicators, namely that their development can push partners to be more explicit and clear in what they are trying to achieve. They also form a verifiable basis for judgements made as part of the evaluation, something that is surely appreciated by those evaluated and funders alike. At the same time, particularly in projects that aspire to achieve social impact, quantitative indicators have been viewed as providing limited value for grasping what is really going on in terms of change.

“And there is a real issue which is that sometimes quantitative indicators are kind of reductive relative to the complexity of interventions.” (Source: OASIS representative)

Yet, despite a certain dose of scepticism towards quantitative indicators, the partners tried to reconcile the need for hard numbers and softer understanding. In Athens [Curing the Limbo](#), indicators were developed while proceeding with the project. This may mean that the measurement will not be as full as it would have been otherwise, but the more specific character of indicators could actually provide a more precise picture of the effects.

“If you think more qualitatively, every time you go back, you think and you reframe. When you think

quantitatively you start with indicators right from the first day. We did a mixed model where we developed our indicators as we were implementing activities, seeing what indicators would be the most reflective of the processes.” (Source: Curing the Limbo representative)

In the Aveiro [Steam City](#) project, the indicators – which form basis of its M&E approach – were developed collaboratively at the beginning of the project but are treated in a dynamic and flexible way. The partners are thus ready to change them should such a need arise based on their insights from project implementation.

“It is an ongoing KPI model. For us it’s very clear that if we have major breakthroughs, or if we have some constraints or limitations, if we find something else, we will change it.” (Source: STEAM City representative)

Some analysed UIA M&E case studies noticed that what is at stake in evaluation really goes beyond indicators and measurement, which warrants a conclusion that different bases for judgement, more qualitative, should ideally be developed. Importantly, the numbers themselves, while perhaps offering more certainty, can only go so far in grasping attention and translating the project achievements into meaningful information for the evaluation’s target audience. In this sense, stakeholder perceptions, testimonies and stories gathered during the evaluation can capture the importance of specific interventions, but also show how and why things really happen.

Importantly, adopting a M&E model based only on indicators, without a wider conceptual analytical model, comes with several risks. Firstly, it is the risk of losing the ‘human dimension’ of the project, and instead focusing on tangible, material effects, which are usually project outputs. By failing to grasp the process of obtaining these effects, the project often becomes a ‘black box’, and understanding the ‘who’, ‘how’ and ‘why’ is difficult. An indicator-based model is often a predefined assessment framework that is not necessarily capable of grasping the dynamics of implementation, responses to challenges or creativity of adopted solutions. Arguably, all of these should be observed in an innovative project.



[See on UIA website](#)

