CEDIL LESSONS LEARNED PAPER 1

The use of middle-level theory in CEDIL-funded research studies

Howard White

About CEDIL

The Centre of Excellence for Development Impact and Learning (CEDIL) is an academic consortium supported by the UK government through UK Aid. The mission of the Centre is to test innovative methodologies in evaluation and evidence synthesis and promote evidence-informed development. CEDIL-supported projects fall into three programmes of work: evaluating complex interventions, enhancing evidence transferability, and increasing evidence use.

CEDIL Lessons Learned Paper

The CEDIL Lessons Learned series shares learning from across the CEDIL programme.

About this working paper

The use of middle-level theory in CEDIL-funded research studies

Authors

Howard White

Institute

Campbell Collaboration

Contents

List of abbreviations

CCA	Causal Chain Analysis
CCT	Conditional Cash Transfer
CEDIL	Centre of Excellence for Development Impact and Learning
EEF	Education Endowment Foundation
FCDO	Foreign, Commonwealth and Development Office
LSHTM	London School of Hygiene and Tropical Medicine
MLT	Middle-level Theory
PoW	Programmes of Work
QCA	Qualitative Comparative Analysis
SAcc	Social Accountability
TaRL	Teaching at the Right Level
ToC	Theory of Change

Abstract

Middle-level theory (MLT) sits between project-level theory, which is specific to a particular context, and grand theory, which is too general to be empirically useful. By understanding the underlying causal processes of a class of interventions, as well as the factors conditioning their operation, MLT seeks to enhance the transferability of study findings and so inform programme selection and design. During its inception stage, enhancing transferability through the use of MLT was identified as a Programme of Work for the UK Aid-funded Centre of Excellence for Development Impact and Learning project, which commissioned 17 studies using such an approach. This paper reviews the experience of these studies and proposes a research agenda for further use of MLT.

1. Background and introduction

The Centre of Excellence for Development Impact and Learning (CEDIL) is a multi-year research programme based around a logical structure for producing, testing, and disseminating innovative approaches to impact evaluation and evidence synthesis.[1] This structure was to produce a series of 'pre-inception' papers on the gaps in evidence and methods that CEDIL was to address, followed by 'inception papers' to further develop the opportunities identified in the pre-inception papers, which would inform CEDIL's programmes of work (PoW) through a set of commissioned studies informed by methods discussed in the inception papers. Finally, lessons learned papers would summarise what we have learned from the commissioned studies about the application of these methods in practice.

This approach has been followed by CEDIL in the case of middle-level theory, which was identified as one of three PoW for commissioned studies, with two exceptions.[2] First, it was agreed to add a set of 'fast-track studies' to further operationalise the ideas laid out in the inception papers. Second, the round of cuts imposed on UK Aid from 2020 meant that approximately one-third of the commissioned studies were discontinued.

The pre-inception paper by a team led by researchers from the London School of Hygiene and Tropical Medicine (LSHTM) – 'Gaps in evaluation methods for addressing challenging contexts in development' (Davey *et al*., 2017) – identified theories of change (ToCs) as an area deserving of more attention in evaluations. The LSHTM team followed this up with the inception paper by Davey *et al*. (2018), titled 'Designing evaluations to provide evidence to inform action in new settings', which picked up on middle-level theory (MLT), laying it out more fully, leading it to be adopted in the PoW. An additional paper on representing ToCs by Rick Davies was commissioned to complement these papers (Davies, 2018).

One purpose of the fast-track papers was to operationalise ideas identified in the inception papers. In this case, the ideas presented by the LSHTM team were elaborated in more practical form in a subsequent methods paper and brief (Cartwright *et al.*, 2020; Cartwright, 2020). As laid out in those publications, MLT informs an understanding of the transferability of study findings and so can be used to inform programme design and evaluation.

Hence, one of the three CEDIL PoW was entitled 'Enhancing transferability through the use of middle-level theory'. Seventeen commissioned studies planned to use middle-level theory. The main purpose of this paper is to summarise this experience and the lessons learned from applying MLT in those studies.

Part 2 describes the evolution of middle-level theory in CEDIL, summarising the papers mentioned above. Parts 3 and 4 present the planned approaches to MLT from the CEDIL projects, drawing on analysis of proposals and inception papers. Specifically, Part 3 discusses how the MLT was developed in the commissioned studies and Part 4 the planned uses of that MLT. Part 5 examines how MLT was actually used in CEDIL papers, while Part 6 sketches out a further research agenda for future use.

---

[1] See White and Masset (2018) for more details on CEDIL.
[2] The other two PoW were evaluating complexity, and use of evidence.

2.   The evolution of middle-level theory in CEDIL

The conceptualisation of MLT in CEDIL

ToCs have become a standard part of programme and evaluation design in international development. ToC in evaluation was put forward by evaluators in the 1970s such as Carol Weiss (e.g. Weiss, 1977). The approach was given further impetus by the development of realist evaluation, with its central focus on context-outcome-mechanism (Pawson and Tilley, 1997; Pawson, 2006).

ToCs have, to a large extent, replaced logframes in international development. Logframes list expected achievements (outputs and outcomes) at different levels, whereas ToCs examine how each step is intended to lead to the next. The difference between a logframe and a ToC is thus the focus of the latter on causal mechanisms, and the assumptions of what needs to be in place (the context) for those causal mechanisms to operate. For example, a school feeding programme may increase attendance at school but will only improve learning outcomes if teachers are not absent. Carvalho and White (2004) present an early application of ToCs in international development, while Vogel (2012) outlines their use more generally.

However, the central premise of this paper is that MLT is different to a ToC, and that using an MLT framework can add value, especially in regard to the transferability of study findings. Whether this has proven to be the case in the work funded by CEDIL is an issue we return to in the final section of this paper.

The idea of MLT has been around since the late 1940s, being first advanced by Robert Merton (1949), and later developed by other sociologists such as Raymond Boudon (1986). Merton argued that trying to derive a general theory for the whole of society was ultimately of no use, since such grand theories were too vague to be empirically useful. On the other hand, purely empirically driven approaches would be too bogged down in the specifics of the context from which the data were collected. Middle-level theory groups together observed empirical regularities, such as a common causal mechanisms.

Unlike ToCs, MLT has not been explicitly used much in international development. Hence, Davey *et al*. (2018) lays out the basic ideas of MLT for a development audience. Like Merton, the authors state that MLT lays between specific project ToCs and a grand theory such as Marx's theory of class stratification or Foucault's theory of governmentality (2018: 2).

For example, Marx's argument that classes are based on relationship to the means of production, and that those with control over the means of production have greater wealth and power, is a grand theory for all societies, which needs greater detail in specific applications to specific societies (modes of production) to see its application. Following such a theory, the notion of the articulation of modes of production – specifically that the form a new mode of production takes in any society is mediated by the pre-existing mode of production – has been applied in specific cases in the work of Goran Hyden in analysing the failure of collectivisation in Tanzania, as peasants could 'retreat' into subsistence farming (Hyden, 1980), and his more general framework of 'the economy of affection' and the challenges an 'uncaptured peasantry' for the development of a market-based economy (Hyden, 1983).

Examples of middle-level theories cited by Davey *et al*. (2018) are (i) cognitive dissonance (when a person's actions do not align with their beliefs, which can cause stress and anxiety);

(ii) educational theories about metacognition (literally 'thinking about thinking', which in practice translates into teaching effective learning strategies); and (iii) behavioural economic theories, which examine how behaviour is influenced by factors beyond simple economic incentives. To elaborate the last of these, people value equity as well as income. This point is illustrated by the dictator game, in which Person A is given an amount of money (e.g. US$ 100). Person A can chose to give any amount of this US$ 100 to Person B. Person B has to accept the gift. If they do not then Person A also gets nothing. 'Rational man' responding to economic incentives in traditional economics would both offer and would accept US$ 1. However, experience shows that people playing the role of Person B typically reject offers that are less than around 20% of the value received by Person A. The conclusion derived from this is that people care about fairness (equity).

Realist evaluators similarly state that their approach has little time for grand theory, needing instead 'middle-range, substantive theory', for which they need to 'ransack' grand theory for 'propositions relating mechanisms, context and outcomes' (Pawson and Tilley, 1997: 138).

As may be clear from these examples, the middle level between grand theory and project-level theory is quite broad. So, where best to be the middle of this wide middle?

Examples of different levels of theory are shown in Figure 1. There are very many theories of behaviour change, with many of them starting with a person's knowledge, perceptions, and attitudes, which need to be changed for behaviour to change. Hence, a large number of interventions to change behaviour provide information: for example, eat more fresh fruit and vegetables because they are good for you; exercise more as it is also good for you; driving when drunk is dangerous; stop smoking because it is bad for you. The example of providing information on the harms of smoking is a specific example, making the theory more specific. It becomes more specific still by stating that the information is provided as messages on cigarette packets, and even more specific by stating visual images should be used. For example, in the UK the messaging has gone from a written message on the side of the packet to stark visual images of the adverse health effects of smoking.

Figure 1 Different levels of theory

General theory of behaviour change

Providing information on the benefits or harms of a behaviour supports adoption or cessation of the behaviour

Providing information on the harms of smoking reduce smoking

Messages regarding health risks on cigarette packets reduce smoking

Visual images regarding health risks on cigarette packets reduce smoking
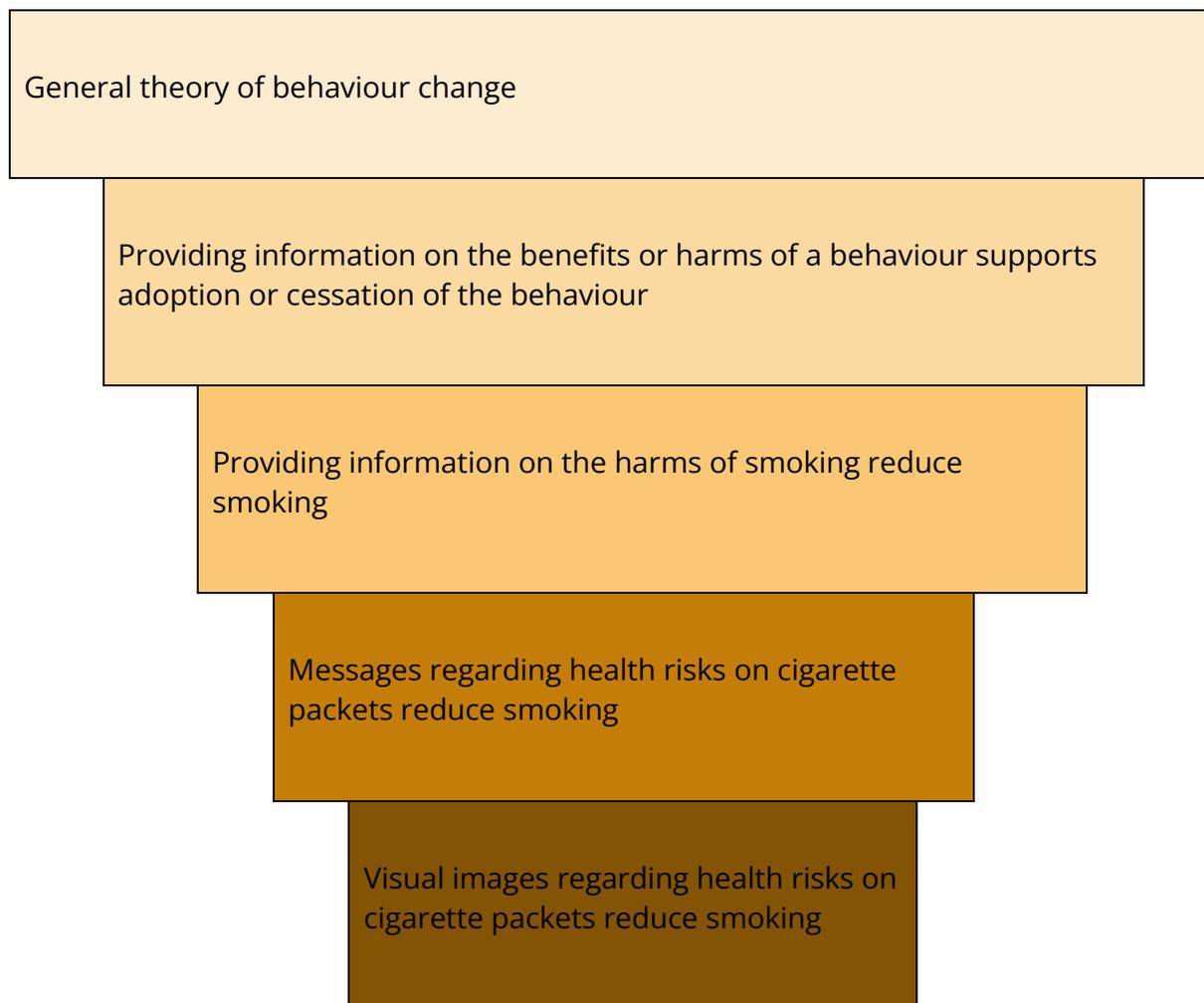
The top level is clearly a grand theory, which needs to be made more specific to apply in any particular case or set of cases. The bottom level is quite specific. In this example, probably the best place to be is the middle row, which encompasses the next two rows, but also other means of providing information, such as TV adverts, school-based information campaigns, and so on. As explained below, these choices preserve fidelity of function but do not retain fidelity to form. While ToCs may well be presented as 'provide information > change in knowledge and attitudes > change in behaviour', such a presentation does not actually dig into the causal mechanism behind these links. I return to this issue below.

The more abstract levels in Figure 1 may also be referred to as 'loose theory' (Davies, 2016). Loose theories can be used when the design of the intervention is not yet fully determined, and outcomes may be emergent. The theory gets thickened (i.e. moves to a more specific level) as the intervention design becomes more fully specified.

Any specific programme theory will draw on multiple middle-level theories, and evaluations can inform refinement of MLTs by testing specific assumptions. Davey *et al.* (2018) identify four evaluation approaches to building MLT:

1 Framing evaluation questions using MLT, so that the evaluation can test causal processes;

2   Using process evaluations and mixed methods, which can help identify contextual factors of importance for causal processes to operate;

3   Leveraging heterogeneity to understand what works in which context and why (across-case analysis of heterogeneity); and

4   Leverage heterogeneity from case studies to delve more deeply into variations in the operation of causal processes (within-case analysis of heterogeneity).

Ideally, this approach will allow the identification of contextual 'markers' that can be used to indicate if a causal process may be expected to operate successfully or not in a particular setting. That is, MLT helps identify settings in which the causal process utilised by an intervention may operate.

This leads to a very useful distinction between generalisability and transferability. The former would mean that a finding is universally valid, which is a very bold claim. But transferability means that the finding can apply in some settings but not others, which seems a reasonable expectation. We rest between the extremes of constructivists who claim everything is unique so no study can lay claim to learning anything that may apply elsewhere and those who would make universal claims based on a single study. The contribution of MLT over ToC as currently practised is partly this explicit attention to external validity, and so to establishing conditions or settings to which a finding (a causal mechanism) may apply. The marker is thus some indication of a relevant aspect of the context in realist terminology.

An example of a marker is provided by IZA's World of Labour evidence maps.[3] The maps are global maps showing which countries the evidence for a topic comes from, with countries divided into one of five types: 1) innovation-driven economy; 2) efficiency-driven economy in transition to a more advanced stage; 3) efficiency-driven economy; 4) factor-driven economy in transition to a more advanced stage; and 5) factor-driven economy. It is expected that the effects of the various labour market interventions examined will be different in these different contexts, as the workings of the underlying causal processes will vary by the type of economy.

A further concept related to transferability is that of transportability, which refers to the transporting of an intervention from one setting to another. Discussion of 'adopt versus adapt' can be found across many literatures: should we simply adopt interventions which have been proven to work elsewhere, or should we be adapting them to our context? Since evidence-based decision making should not be seen as a blueprint approach, this would seem to suggest a preference for adapt and test, or possibly adopt, test, adapt.

The use of MLT allows for a more refined version of adopt versus adapt. Davey *et al*. (2018) suggest that in transferring an intervention we should wish to preserve fidelity of function rather than fidelity of form. The latter would imply adopting the specifics of the intervention, whereas the former suggests that we should transfer the causal principle (that is the specific bit of theory underlying the observed causal mechanism), but the specific form of the intervention may vary. So, we may take the causal principles of providing information on the benefits of a practice (function), but the specific form of media used to communicate the information will vary by context. What we often see in practice is the opposite: the form is

adopted but not the function. Borrowing a term from evolutionary biology, this has been labelled 'isomorphic mimicry' (see Andrews *et al*., 2013).

Davey *et al*. (2018) provide some examples where there has and has not been transferability. Parenting classes to improve child health and development outcomes have been proven to transfer from developed to developing country settings. Thus, the causal process of parents acquiring and successfully adopting desired parenting behaviours, which they did not already practice, appears to hold across these settings. In contrast, an HIV communications programme that had worked in New York did not work in Scotland because it did not take place until some years later, by which time awareness of HIV/AIDS was much more widespread. For further examples based on road safety, see Box 1.

Box 1: Are findings from road safety studies transferable?

The most striking finding from a map of road safety interventions is the geographical distribution of evidence. While over 90% of traffic fatalities are in low- and middle-income countries, over 90% of the available evidence comes from North America. Does this massive inequity in the source of evidence matter? That is, what about transferability? Do study findings from North America apply in developing countries? The answer is yes and no.

On the yes side, some things are obviously universal. When it is dark people cannot see as well, and that makes driving more dangerous; thus, lighting is effective wherever you are. Drinking alcohol impairs your judgement and slows your reaction time and so greatly increases the chances of an accident. So, cutting back on drink driving will save lives. And it is well established that speeding is linked to accidents. So, requiring speed limiters in vehicles, as is being done in Rwanda, will reduce the number of accidents.

It is also well established that driver education and public safety campaigns are ineffective. The reason that people speed or drive through red lights is not that they did not learn that red means stop. Of course, they know this. But people have a poor intuitive understanding of risk, and a feeling that the increased risks from dangerous driving do not apply to them. Moreover, the majority of people think that they are a better-than-average driver. Less than 3% think they are worse than average. Of these, nearly all are women, who are actually, on average, safer drivers than men.

But there are other areas in which the research imbalance matters as the research is not transferable, leaving large evidence gaps. In developed countries, cars are the main killers and the majority of people killed are car passengers or drivers. However, in developing countries most people killed are pedestrians or cyclists. Furthermore, cars account for only 20% of road deaths; in developing countries, most people are killed by buses and trucks. But the empirical studies on front-end engineering – to reduce the chances of injury and death when a vehicle hits a person – are focused on cars. Moreover, there is a whole class of indigenous vehicles, such as tuk tuks, rickshaws, and jeepneys, for which there are no empirical studies on safety, again despite a body of literature drawing attention to them. Finally, there is the issue of different road use patterns, whether it is cows wandering across highways in India or the ubiquitous small shops, street sellers, tea stalls, and restaurants alongside major roads in many other developing countries.

The common element on both the yes and the no side in this discussion is that the argument is informed by ToCs. It is the statement of the theory that makes it clear whether there is transferability of study findings or not. So, 'people can see less well in the dark, and when people can see less well they are more likely to have an accident' is a simple theory that is clearly universal and so transferable. On the other hand, the theory that 'improving front-end engineering for cars will reduce traffic fatalities' is far less transferable, since only one-fifth of traffic deaths in developing countries are caused by cars.

Source: White, H. (2020) 'When, where and for whom does evidence apply? The example of road safety'. Campbell Collaboration Blog, published 21 February 2020. www.campbellcollaboration.org/blog/road-safety.html

Applying MLT to programme design

Once the inception stage was completed, the CEDIL Secretariat and the UK Department for International Development (now the Foreign, Commonwealth and Development Office (FCDO)) agreed it would be useful to commission additional papers to further develop some of the ideas in the inception papers. One such area of interest was middle-level theory. Originally, it was hoped that a number of papers on this subject would be commissioned. In the event, however, just one paper was; that by Nancy Cartwright and colleagues.

The methods paper by Cartwright *et al*. (2020) elaborates how to develop an MLT, using the example of a conditional cash transfer (CCT). A middle-level theory is 'thickened' to move down to a project-level theory. The different levels in Figure 1 illustrate such a thickening. As part of their exposition, the authors distinguish different types of 'assumption', which are explained below.

The accompanying methods brief (Cartwright, 2020) lays out 10 steps she proposed are involved in the development of an MLT:

1. **Specify the overall middle-level ToC**: this is a statement of the main underlying theory behind an intervention, e.g. 'providing a cash incentive for a behaviour will encourage that behaviour'.

2. **Produce a step-by-step diagram**: ToCs are typically presented in diagrammatic form with arrows, which is the approach suggested here. Most ToCs list inputs, activities, outputs, and outcomes that are linked by arrows.

3. **Describe the causal principles at work at each stage**. The arrows in ToC diagrams represent the causal processes. As Davies (2018) has argued in another CEDIL paper, these arrows need more attention as the nature of the process may vary (e.g. linear or non-linear, a probabilistic relationship or a necessary condition, unidirectional or bidirectional, etc.). Cartwright urges us to spell out more carefully the causal mechanism. She labels the causal principle underlying the causal mechanism as one of the assumptions in MLT. While calling the causal principle an assumption is not universal or even common, it does highlight that we should not assume it is necessarily true, but that it should hold for the intervention to work as intended. Of course, there may be other necessary conditions for it operate, or there may be moderators that affect the strength of the mechanism – these are other assumptions that we come to below.

   Any middle-level theory will have several causal processes. So, in the case of CCTs, we also assume that schools will accept the children and that they will learn something should they attend school.

4. **Add support factors to the diagram**. Support factors are things (which may be part of the intervention design, but not necessarily so) that support the assumed causal process to operate. Thus, providing clear information to parents as to the conditions, and monitoring adherence to those conditions, makes it more likely the CCT will have the desired effect of increasing school attendance. The presence of support factors is another assumption.

5. **Add derailers to the diagram**. Derailers are factors that may prevent a causal process from operating as intended. For example, teacher absenteeism will inhibit children from learning even if they do attend school.

6. **Add safeguards against the derailers to the diagram**. Safeguards limit the disruptive effects of derailers. CCTs are a demand-side intervention. They may work best if accompanied by supply interventions that ensure sufficiency of the quantity and quality of the services for which demand is being supported by the cash transfer.

7. **Allow for causal loops**. A step in the theory may loop back to an earlier step. If parents were not previously sending their child to school, they will gain more information about schooling once they do so. This may encourage them to send their other children to school – or discourage them from doing so, depending on the schooling experience.

8. **Specify the expected range of application**. Interventions will not work in all settings. If primary education is already universal then a CCT with the condition to send children to primary school will have no effect on school enrolment. If school facilities are not available, or are known to be of poor quality, or as havens of abuse, then a cash transfer will not help. Similarly, if there are cultural beliefs that inhibit the schooling of girls then cash alone is unlikely to be sufficient to change that.

9. **Draw out the implications for monitoring and for evaluation questions and indicators**. All of the above points have implications for what the monitoring and evaluation system should try to capture. Programme management needs information on compliance with conditions. Evaluations may focus on the upper end of the MLT, that is the success or otherwise of improved education outcomes.

10. **Draw out the implications for future programme design**. What sort of safeguards are needed in what contexts to increase the probability of programme success? As mentioned above, a complementary supply-side intervention may assist the success of a CCT.

3.   How are studies developing their middle-level theory?

CEDIL has three research programmes – or PoW. These PoW are: (PoW1) evaluating complexity; (PoW2) enhancing transferability through the use of middle-level theory; and (PoW3) supporting evidence use. This section provides an overview of how CEDIL-funded studies planned to apply middle-level theory.[4]

The inclusion criteria for this paper are that the study is CEDIL-funded under the PoW and that it explicitly adopts middle-level theory. This includes all studies funded under the MLT-focused PoW2, as well as studies from the PoW1 and PoW3 that apply middle-level theory. The studies are listed in Annex 1. This list includes the studies that were later discontinued as a result of cuts in the UK Aid budget in 2020.

The paper first discusses the different approaches to developing a middle-level theory, and then how MLT is being used. The final part discusses a possible research agenda.

Approaches to developing middle-level theory

There are three approaches to developing middle-level theories: (1) the bottom-up approach; (2) the top-down approach; and (3) an iterative approach. The bottom-up approach either (i) aggregates across project-level ToCs to develop a middle-level theory or (ii) is data driven, with the theory emerging from analysis of either quantitative or qualitative data from intervention studies. The top-down approach starts out with the middle-level theory and then uses the evidence from evaluations of different interventions to either (i) test the theory or (ii) evaluate the interventions. I provide more discussion of these differing approaches below. The third approach falls between these two, and involves developing the middle-level theory in an iterative manner going between theory and data. Of course, in practice, the iterative approach has to start either with data (bottom up) or theory (top down). Likewise, both top-down and bottom-up approaches will usually adopt some measure of iteration. Nonetheless, each CEDIL study can be fitted into one of these three categories, as shown in Table 1.

Table 1 shows the three categories as columns. It also shows rows indicating whether a study is qualitative, formalised qualitative, or quantitative (modelling and statistical). The final category of studies may be top down if they specify a mathematical model that is then parameterised using data from various evaluations. However, the modelling approach may be more bottom up if it is data driven, such as adopting a machine-learning approach to model formulation. As stated above, in practice all approaches will contain some iteration – some more explicitly than others. Formalised qualitative refers to studies that code intervention design and contextual characteristics to be subject to some formal analysis, such as qualitative comparative analysis (QCA).

---

[4] This section was prepared as an *ex ante* lessons learned paper for internal use to facilitate cross-learning between studies. The contents were presented at a workshop for study teams in February 2021.

Table 1: Approaches to developing middle-level theory

	Bottom up or data driven	Top down or model based	Iterative
Qualitative	Scaling social accountability (SAcc) for health Catalysing responsive and inclusive governance Finance for supporting climate change adaptation	Language transitioning research synthesis	Gender and social outcomes of water, sanitation, and hygiene (WASH) interventions
Formalised qualitative	Demand for contraception among adolescents	Involving men and boys in family planning	The art and science of using evidence
Quantitative	Youth skills training programmes Integrating data and theory Teaching at the Right Level's (TaRL) effectiveness and generalisability	SHARPE project in Ethiopia Transferability of education Predicting optimal policies Syrian business development Structural estimation of spatial spillover effects of cash transfers	An empirically driven theory of poverty reduction

An example of the bottom-up approach is the study of SAcc in health, for which the stated intention is to 'develop a middle-level theoretical framework to evaluate how and under what contextual conditions development outcomes are improved where scaled-up social accountability is incorporated as a sustainable part of development policymaking'. The authors proposed to 'focus on unpacking the value added of SAcc for broader health sector reforms as well as look into how and why SAcc is incorporated into decision making, policymaking, and implementation – with careful assessment of how and why the champions of SAcc were able to win support for its scaled-up inclusion. This project shifts the focus of SAcc research and evaluation. *It looks beyond results of individual projects, in terms of information provision, to why and how they are integrated within the political economy of policymaking processes* in health and to what effects' (italics mine).

An example of the top-down approach is the systematic review of involving men and boys in family planning. In their design paper the study team present an elaborated ToC, stating that

they will test and adapt the logic of the model: 'Working with our international expert advisory group and using a Causal Chain Analysis (CCA) approach we will test and adapt the logic model presented in Figure 1'.

In the case of the men and boys study, despite having a fully articulated model, there is an intention to iterate between the model (theory) and the data. A more explicit example of the iterative approach is the study on achieving gender and social inclusion through WASH interventions. In this case, the team say very explicitly that they intend to develop the ToC through an iterative process: 'the logic model will be built iteratively, informed by theory and refined by stakeholders... we will discuss the complementarity, synergies and divergencies between the two lines of enquiry. As a result of this process, the logic model will be expanded and adjusted.' The WASH study team state they will use framework synthesis, i.e. a systematic review in which coding analysis uses a pre-specified conceptual framework. MLT is well suited to providing such a framework, and well suited to reviews of development interventions in which studies are likely to take place in a range of settings.

Qualitative versus quantitative analysis

The purpose of CEDIL is to explore innovative approaches to impact evaluation. The focus of PoW2 is not so much on the evaluation methodology, but rather the development and use of middle-level theory as a framework for designing, undertaking, and using impact evaluations.

A middle-level theory may be seen as a structural model, which may be estimated (parameterised) by using the data from intervention studies. The quantitative studies generally use statistical testing. In one case, theory testing is formalised by using out-of-sample model validation; that is, by developing the theory using data from a set of interventions and then testing it against data from another intervention. A good example of the mathematical approach is the project 'Building optimal policies', in which the research team take two approaches to estimating a structural model embodying a middle-level theory. Studies that fit the data to an existing model are top-down approaches. More data-driven approaches, such as the studies using machine-learning approaches and those using meta-regression, adopt a bottom-up approach.

Some of the qualitative studies proposed using QCA to identify associated combinations of context, mechanism, and outcomes. One study (SAcc in health) explicitly states that it will use process tracing. Generally, however, one issue raised by the studies is how qualitative theories can contribute to causal analysis in the context of a middle-level theory approach.

4.   What is the relationship between the study and middle-level theory?

Studies may engage with middle-level theory in one of four ways: (1) Use the study to develop a middle-level theory; (2) Use the MLT to test interventions, i.e. to assess why certain interventions work or did not work; (3) Use evidence from intervention studies to test or validate the middle-level theory; and (4) Use theory to make predictions about the expected success or failure of an intervention (or adaptations of an existing intervention) in a new setting or context, or to make predictions about a new intervention in the same context. The last of these brings us closest to the intention of this PoW: that is, middle-level theory can be used to inform the design of programmes in order to increase the likelihood of their success. This is the approach in the CEDIL Methods Paper by Nancy Cartwright and colleagues (Cartwright *et al.*, 2020).

Table 2 on the next page classifies the studies under these four headings. Some studies appear in more than one column.

The majority of the studies planned to develop the middle-level theory as part of the study. In a small number of cases, the MLT is an end point of the study with no further use of it. For example, the proposal for the study of youth employment programmes in Uganda states that the team planned to '(1) apply regression tree ([machine learning], ML) methods to study Heterogenous Treatment Effects (HTEs) to inform the generalizability of the observed impacts, and (2) expand on ML methods to conduct causal mediation analysis. These methodological aspects will contribute to the formulation of middle-level theories'.

Table 2: Uses of middle-level theory in CEDIL-funded studies

Develop theory	Test interventions	Test theory	Predict impact
Demand for contraception among adolescents	Language transitioning research synthesis	The art and science of using evidence (unclear)	Transferability of education
Scaling SAcc for health	Involving men and boys in family planning	Involving men and boys in family planning	TaRL's effectiveness and generalisability
Catalysing responsive and inclusive governance	SHARPE project in Ethiopia	Predicting optimal policies	Structural estimation of spatial spillover effects of cash transfers
Finance for supporting climate change adaptation	Demand for contraception among adolescents	Gender and social outcomes of WASH interventions	
Language transitioning research synthesis	The art and science of using evidence		
Gender and social outcomes of WASH interventions	Finance for supporting climate change adaptation		
An empirically driven theory of poverty reduction			
Youth skills training programmes			
Integrating data and theory			
Syrian business development			

Several of the mathematical studies use machine learning. It might be asked if machine learning is not inherently atheoretic in its starting point, given that it involves a data-driven approach to developing models. The usual response is that nothing can be completely atheoretic, as theory informs variables selection for the analysis. But as Big Data approaches make a growing number of variables available, this response is still partly true but less binding. In practice, CEDIL-funded studies adopting mathematical approaches rely on differing degrees of iteration between theory and data. For example, the study to predict optimal policies is based on a specified social welfare function to which the data are fitted.

In most cases the studies have both presented or developed the theory, which is then applied in another way such as theory testing. For example: 'we will test and adapt the logic model' (men and boys in family planning) and for the analysis of the social inclusion effects of WASH programmes: 'the logic model will be built iteratively, informed by theory and refined by stakeholders'. The second example raises a question as to what constitutes evidence in the

iterative development of a middle-level theory. Is stakeholder engagement a step in theory development, or does it count as theory testing?

Studies may either take the theory as correct and use it to test interventions or use evidence from evaluations to test the theory. Furthermore, that theory testing might be quantitative or qualitative. In the CEDIL-funded studies, testing interventions is usually cast in terms of analysing whether particular programme components are effective or why the intervention did not work in certain settings. For example, AIR's review of language transitioning research the language of instruction seeks to 'explain why certain language of instruction policies are likely to be more effective than others', although the paper does not explicitly refer to MLT.

Only a minority of studies explicitly plan to use middle-level theory to examine how interventions will work in another setting, i.e. predicting impact. For example, the study by the Educational Endowment Foundation examines if evidence-based teaching approaches being applied in the UK can be transferred to a West African setting. Specifically, it uses meta-regressions with a pre-specified set of coded study features with moderator variable analysis to answer questions about the transferability of evidence.

5. Use of MLT in CEDIL papers in practice

Parts 3 and 4 discussed the planned development and use of MLT in the CEDIL papers. What did the studies do in practice?

Four different applications of MLT can be identified in CEDIL papers: (i) Development of MLT, which is in virtually all papers, though a few have no explicit theory; (ii) Testing interventions in a theory-informed way; (iii) Testing theory or, more specifically, hypotheses derived from the theory; and (iv) Informing programme design by predicting design features with greater impact.

Developing middle-level theory

The studies used four approaches to developing mid-level theory: (i) Drawing on existing conceptual frameworks; (ii) Review of existing literature; (iii) Theoretically informed; and (iv) Empirically based. These approaches are not mutually exclusive, with most studies using more than one of these approaches (see Table 2).

Examples of the different approaches taken to developing middle-level theory were:

- *Existing conceptual frameworks*: for example, the authors of the review on Structural interventions aiming to enable adolescent use of contraception (Burchett et al., 2022) state that they build on an existing conceptual framework by the International Centre for Research on Women ((Glinski et al., 2014).

- *Review of existing literature:* The study of language of instruction reviews the existing literature to construct a ToC that informs construction of a programme typology and moderators of effectiveness (Nakamura *et al*., 2022). Since the literature review is of theory this study also falls under theoretically informed MLTs.

- *Theoretically informed:* The study of heterogeneity of impact of cash transfers is mainly empirically driven, but initial variable selection is also informed by theory (Handa *et al*., 2023). The authors state that their intention is to develop a theory as to which households will benefit most from cash transfers.

- *Empirically based*: As indicated in the project name, for 'an empirically driven theory of poverty reduction' the ToC is derived from empirical patterns identified by using machine learning to examine the characteristics of those for whom unconditional cash transfers have high and low impact, although the initial variable selection was informed by theory. As the authors say: 'putting together these different pieces of information (pre-treatment characteristics and post-treatment behaviours) can help us understand the different pathways out of poverty, and ultimately contribute to a middle-level theory of sustained poverty reduction' (Handa *et al*., 2023).

Test interventions

Interventions are tested in the CEDIL studies in both reviews and primary studies. The most common approach is to use theory to generate hypotheses regarding moderators related to context, population, or programme design or implementation. Thus, studies seek to exploit

heterogeneity to determine where or for who interventions are most effective, rather than blanket statements of an intervention working or not. This approach of course speaks to the notion of transferability resting on the idea of understanding the contextual factors ("the setting") that condition the size of programme effects, which are ideally informed by an understanding of the conditions required for a causal process to operate.

Many of these studies also test causal processes, illustrating the point that studies need not be restricted to just one of evaluating interventions or testing causal processes. For example, the study by the Education Endowment Foundation (EEF) and eBase conducted reviews of various education interventions, with the intention of testing transferability of interventions by testing country income level as a moderator. However, there were too few studies from developing countries for this analysis to be possible (e.g. on the use of corporal punishment in schools).[5]

Several of the reviews develop a middle-level theory to generate hypotheses that can be tested through further intervention studies. In one case, TaRL, the CEDIL-funded study supported both the review and a follow-on primary study to confirm one of the hypotheses from the review (Angrist and Meager, 2022). This is elaborated on in section 5.4.

Test theory

The study on scaling up SAcc interventions identified three approaches or pathways to scaling up: resonance (build on expertise and existing experience within government), resistance (leveraging conflict by mobilising civil society to advocate for change), and best practice (a technocratic approach based on evidence of what is working elsewhere). The first pathway sees change happen chiefly as a response to new information, the second sees change happen chiefly through conflict, and the third through deliberation, compromise, and collective action. The authors develop and test these theories based on existing literature – and more such case study work is planned. Based on their work thus far, while saying that a blend of the three approaches could be used, they suggest that resonance is likely underutilised. They also find that there is a temporal dimension in many cases, where actors change path from one pathway to another (in response to internal and external moderating conditions). They point out that the research synthesis using QCA by Fox *et al*. (2022) also demonstrates this temporal dimension empirically to some degree, although Fox et al. mainly use resistance for their theoretical framing.

The review of structural family planning interventions used a conceptual framework that suggests that interventions need to be adapted to the life stage of the target group. However, they found that the included studies rarely specified the life stage, which meant that the theory could not be tested (Aventin et al., 2023).

---

[5] Since the use of corporal punishment is widespread, even where prohibited, it would be useful to have evidence of its (in)effectiveness. I am not proposing that we randomly assign schools to a regime of corporal punishment. Rather, a 'discouragement' design could be used. That is, rather than an encouragement design in which an encouragement to engage in the intervention is randomly assigned, in this case a discouragement from use of corporal punishment would be assigned at random.

Inform programme design and predict impact

Middle-level theory can be used to make predictions as to the impact of different interventions and intervention designs. This can take the form of recommendations of design features expected to yield higher impact. For example, the review of structural interventions for family planning concludes that the following three steps should be followed: '1) tailor interventions to the adolescent life stages; 2) assess the baseline situation; and 3) select appropriate intervention activities to match gaps, particularly relating to interventions aimed at increasing the desire to limit/avoid/space/delay childbearing, at increasing agency to use contraception and at fostering an enabling environment' (Buchett *et al*., 2023: 4).

In principle, though it is not stated explicitly, Handa *et al*., which identifies those who do and do not benefit from cash transfers based on baseline characteristics, might be used to improve targeting in order to maximise impact

This approach played out most fully in the study of TaRL, which involved a theory-based review to identify factors associated with larger effect sizes. The team then conducted a new randomised controlled trial in Botswana to test the difference one factor (implementation fidelity) made to programme effects, confirming the finding of its importance (Angrist and Meager, 2023). A second example is the study which identifies the pre-treatment characteristics of high flyers, i.e. those who are most likely to utilise cash transfers to escape poverty, which can inform programme targeting to have the largest impact (Handa *et al*., 2022).

Addressing external validity and transferability

The projects vary in the extent to which they address external validity and the range of settings to which they are transferable. For example, the analysis of high flyers – the households who benefit most from a cash transfer – would be valid for a scale-up of the programme in the same setting. However, its atheoretic approach means that the approach does not identify markers to assess transferability. In contrast, the EEF/eBase evidence transfer from UK to West Africa planned to use country income as a marker, and so potentially would test applicability to a wide range of settings. However, the team did not have the data to put this into practice.

The study of TaRL identified a moderator for effectiveness to transfer the same intervention to different settings, although all are in sub-Saharan Africa. The specific finding related to fidelity of implementation, but stated very specifically as 'ensuring children were assigned to the appropriate teaching level'. However, the more general statement would be that 'implementation fidelity matters for effectiveness'. This is undoubtedly true, but is also rather too general to be useful. A more specific statement of what aspect of implementation matters is more useful to practitioners – as indeed was given in this case.

The middle-level theories proposed for social accountability interventions – resonance, resistance, and best practice – are stated at a sufficiently general level to in principle be applicable in all settings. Elaboration of the theories, which is planned for future work, can help identify markers for determining which approach is most appropriate in which setting.

6. Revisiting the MLT research agenda

When assessing the planned research for CEDIL as the research programme got under way, I wrote that: "It was not expected that the CEDIL research programme will come out with The Answer as to how to design and use middle-level theory. Rather it was expected that at the end of the commissioned studies set of studies we are likely to be in the same position as now in neither prescribing nor proscribing any one approach. Rather the CEDIL PoW will develop our understanding of how MLT may be used, and so further the adoption of the approach in appropriate ways to the extent that it improves research or its application to policy and practice". This particular approach was consistent with CEDIL's overall approach, in which the commissioned studies use recent methods so that we learn from how these methods can be used.

But, at the outset, I identified a set of research questions that are repeated here so as to reflect on what we have learned after each question:

1. Can we establish some shared or common language so that research teams are using the same terms to refer to the same concepts? As a related point, can we be clear where the middle is, noting that both Nancy Cartwright and I have said explicitly that there are different levels of middle which rest between local specificity and high theory and we should not get too hung up about it.

At the most basic level, various terms are used: mid and middle, level and range. In my view, this really does not matter and we have settled on middle-level theory here.

Nearly all the papers included in this "lessons learned" paper do refer to MLT. Thus, this new concept has become more widely known about and used as a result of CEDIL. The papers from the inception stage did propose language that could be used, notably fidelity of function and of form, but that was not picked up in the commissioned studies. The paper and brief of Cartwright *et al*. (2018) introduced further terminology in their typology of assumptions, notably support factors, derailers, and safeguards. However, the authors of the commissioned studies were not directed to adopt this terminology and most did not do so. The exception was the SAcc study, for which one of the authors explicitly situated the Cartwright *et al*. (2018) proposed approach in the context of existing approaches to realist evaluation and ToCs.[6]

Hence, thus far, beyond that a wide range of approaches is used. The papers largelty do not adopt a common language, and only a few of the studies referring back to CEDIL conceptual papers on MLT. However, we are aware that there was some influence from the CEDIL events attended by researchers from the study teams on how methods were used. It is expected that MLT will continue to attract attention on account of increased interest in external validity and transferability.

2. How does middle-level theory differ from a good ToC, and what is the value added of the middle-level aspect?

---

[6] https://thomasmtaston.medium.com/pyramids-ladders-and-traveling-theories-3097b39cbe01

In reality, many ToCs for projects are written at a level that can already be seen to sit in the middle-level. My own studies of the Bangladesh Integrated Nutrition Programme (White, 2005; White and Masset, 2006; White, 2010) state assumptions at a fairly general level as 'the programme has to be delivered to the intended target population' and 'key decision-makers affecting a behaviour need to be targeted for behaviour change interventions'. The latter can be stated more specifically as 'fathers and grandmothers (i.e. the mother-in-law of the mother) need to be involved in communications regarding infant feeding practices', which is an example of thickening to the specific context.

From my own perspective, Cartwright *et al*.'s unpacking of categories of assumption – causal process, supports, derailers, and safeguards – has been useful in further understanding the development of ToCs. The emphasis on understanding the causal principle underlying the causal mechanism is important. However, the commissioned studies mostly did not examine that. We are mostly stuck at doing 'X increased the likelihood of Y' without really understanding the underlying causal processes. Further studies should push harder on understanding how things work.

Causal mechanisms

While some studies did examine moderators and use this information to inform programme design. As just stated, I feel we can benefit more from thinking more deeply about the causal process. Let us consider a very common causal link in development interventions: 'provide training > behaviour changes > outcomes improve'. Nearly all development interventions involve some training either of programme staff or intended beneficiaries. So, how does 'provide training > behaviour changes' work? It is commonly assumed that we are providing information people do not have, such that they will then act on that new knowledge. However, that is often not true. People know they should exercise more and eat better. People know that drink driving and speeding are dangerous and that they should stop at red lights, just as they know smoking is bad for their health. Clearly, then, a 'knowledge deficit' causal mechanism may apply in some cases, but not all. If not that, then what? One possible role is exhortation. Another may be peer pressure – a smoker's friends and family may use the graphic pictures on a cigarette packet as a starting point for a discussion. Similarly, group training may create a collective will to act or a 'competition effect', so the causal mechanism is not related to the information at all. Understanding what the causal mechanism is matters, as it has implications for intervention design. For example, in my presentations of MLT), I apply a transtheoretic framework to WASH interventions that suggests that messaging needs to differ for people at the pre-contemplation stage (message: germs are bad) and to those at the planning stage (message: how and when to wash your hands).

Transferability

What did distinguish the CEDIL-funded studies was that the CEDIL PoW had an explicit focus on using MLT to enhance transferability. While that focus is not new, most impact evaluations do not pay much attention to external validity. MLT provides a basis for doing so. This has been an explicit feature of several CEDIL studies, such as the TaRL study already mentioned, in which implementation fidelity was the marker for effectiveness. The EEF/eBase planned to use income *per capita* as a marker, but they did not have the data to do so. These studies

differ from most realist evaluations, which are oriented toward qualitative analysis, thus illustrating how quantitative analysis can be applied to the realist context-mechanism-outcome approach.

3. What is the role of stakeholders in developing middle-level theory? Do their views constitute evidence or are a contribution to developing a MLT to be tested? Can the research programme identify the value of stakeholder-based MLT compared to researcher-driven MLT?

While stakeholder-based approaches were used – e.g. the study on involving men and boys in family planning consulted experts on the study's conceptual framework – there has been no explicit attention to this question in the commissioned studies.[7] This is therefore an area for further research.

4. Where the MLT is being developed bottom up by aggregating across ToCs or programme experiences, how many programmes are needed to do this?

None of the CEDIL-funded studies aggregated across ToCs. One study based its MLT on the adaptation of just one existing conceptual framework. This is arguably a top-down approach rather than a bottom-up one, but can be a recommended approach if there is a widely accepted existing framework.

Given the absence of the bottom-up approach in CEDIL-funded studies, I mention two other studies that have used this approach. The first is an ongoing study of interventions to support financial inclusion. In that case, we identified eight conceptual frameworks by a separate search rather than from the included studies. This approach worked well, and did not take long to do. The other case was another ongoing study for youth employment interventions, where we synthesised across project-level ToCs for different categories of intervention. For this approach it was observed that the principle of saturation set in after 6–8 studies, i.e. no new concepts (causal processes or barriers or facilitators) were identified from coding additional studies. However, few studies explicitly identified the causal process by which the intervention was intended to work. Many ToCs were also weak in identifying the assumptions necessary for the ToC to work. While the evaluations did often identify factors causing weak or missing links in the causal chain, they were not presented as such. No studies presented a revised ToC, updating the *ex ante* ToC in light of the evidence.

Overall, we can draw three conclusions from this. First, conceptual frameworks for interventions are likely to be akin to MLT, and so can often readily be adapted for that purpose. Second, the saturation principle will apply when aggregating ToCs bottom up, such that only 6–8 ToCs are likely to be needed. Third, our experience in the youth employment study showed that many of the assumptions (e.g. that training providing the required skills) did not appear in the ToC but only in the findings.

---

[7] There are two CEDIL papers on stakeholder engagement: Oliver *et al*. (2018) and Oliver *et al*. (2021).

5. Are structural models inherently a middle-level theory? How may machine learning be used to develop such models in a data-driven manner? How does iteration between theory and data enter into machine-learning approaches?

The CEDIL papers include two examples of machine learning being used to identify factors associated with higher impact. The data do not themselves produce the theory as to why these factors have higher impact, but the researchers may fit a theory consistent with the factors that have been identified. As already noted, there is at least an implied theory in the selection of variables on which to collect data or include in the analysis. This is an example of what has been called above a 'loose theory', which could be made more specific.

The two papers had limited iterations. Initial variable selection was informed by theory, and then theory can be developed or refined based on the empirical findings. But neither of the papers finished with a well-developed theory.

6. Can qualitative approaches to middle-level theory formalise the conditions for transferability?

This was not assessed.

7. Final word and next steps

In conclusion, we can say that the CEDIL approach of supporting a number of studies that would use MLT to identify and examine different approaches was successful. A range of approaches to both developing and using MLT in primary studies and reviews was adopted. From this set of studies, we can say that MLT showed promise during the inception phase, and that that promise has been partly realised, though only partly. However, more remains to be done in further research on using MLT, as well as devoting some resources to supporting its use.

Research agenda

- Develop further consensus around terminology and approach. While terminology may seem to be, literally, a semantic issue, agreeing terms can help frame the approach.

- A further programme of studies – both primary studies and systematic reviews – in a specific sector for which a middle-level theory can be developed and tested.

- Examine different approaches to developing MLT, including the role of stakeholders.

- Use MLT to develop evidence-based making products such as guidance, and test these products in practice.

Use

The utility of the MLT approach for development interventions has been demonstrated by the CEDIL programme, but further work is needed to promote the use of MLT. This could be done in the following ways:

1. FCDO creating an MLT resource page, which can include relevant CEDIL resources,[8] as well as a longer reading list. However, it is likely that few would use such material. The CEDIL Methods Brief could be highlighted as the key resource, although that is just one of the approaches to MLT (it is top down, and with little stakeholder engagement).

2. FCDO commissioning a standalone guide on MLT.

3. FCDO requesting MLT be used in some future evaluations or reviews.

4. Further dissemination to the evaluation community.

---

[8] Beyond the end of the CEDIL programme, all CEDIL papers and briefs can be found in Policy Commons (https://policycommons.net), a large, searchable database of high-quality working papers, policy briefs and other grey literature.

References

Angrist, N. and Meager, R. (2022). The role of implementation in generalisability: A synthesis of evidence on targeted educational instruction and a new randomised trial, CEDIL Syntheses Working Paper 4, CEDIL, Oxford. Available at: https://doi.org/10.51744/CSWP4

Andrews, M., Pritchett, L., and Woolcock, M. (2013) 'Escaping capability traps through problem driven iterative adaptation (PDIA)'. *World Development* 51(11), pp. 234–44. Available at: https://doi.org/10.1016/j.worlddev.2013.05.011

Aventin, Á., Robinson, M., Hanratty, J., Keenan, C., Hamilton, J., McAteer, E. R., Tomlinson, M., Clarke, M., Okonofua, F., Bonell, C., & Lohan, M. (2023). Involving men and boys in family planning: A systematic review of the effective components and characteristics of complex interventions in low- and middle-income countries. Campbell Systematic Reviews, 19, e1296. Available at: https://doi.org/10.1002/cl2.1296

Burchett, H., Kneale, D., Griffin, S., Joaquim Picardo, J., de Melo, M. and French, R. (2022). Structural interventions aiming to enable adolescent use of contraception in low- and middle-income countries – final project report, CEDIL Syntheses Working Paper 5, CEDIL, Oxford. Available at https://doi.org/10.51744/CSWP5

Cartwright, N., Charlton, L., Juden, M., Munslow, T., and Williams, R.B. (2020) 'Making predictions of programme success more reliable'. *CEDIL Methods Working Paper 1*. Oxford: CEDIL. Available at: https://doi.org/10.51744/CMWP1

Cartwright, N. (2020). 'Using middle-level theory to improve programme and evaluation design'. *CEDIL Methods Brief*. Oxford: CEDIL. Available at: https://doi.org/10.51744/CMB1

Davey, C., Hassan, S., Bonell, C., Cartwright, N., Humphreys, M., Prost, A., Hargreaves, J. (2017) Gaps in Evaluation Methods for Addressing Challenging Contexts in Development. CEDIL Pre-Inception Paper. London: CEDIL. Available at: https://doi.org/10.51744/CPIP4

Davey, C., Hargreaves, J., Hassan, S., Cartwright, N., Humphreys, M., Masset, E., Prost, A., Gough, D., Oliver, S., Bonell, C. (2018) Designing Evaluations to Provide Evidence to Inform Action in New Settings, CEDIL Inception Paper No 2. London: CEDIL. Available at: https://doi.org/10.51744/CIP2

Davies, R. (2016) 'Evaluating the impact of flexible development interventions using a "loose" theory of change'. A Methods Lab publication. London: Overseas Development Institute.

Davies, R. (2018) 'Representing Theories of Change: A Technical Challenge with Evaluation Consequences'. *CEDIL Inception Paper 15*. Oxford: CEDIL. Available at: https://doi.org/10.51744/CIP15

Glinski, A., M. Sexton and S. Petroni (2014) Understanding the Adolescent Family Planning Evidence Base. International Center for Research on Women (ICRW).

Handa, S., Liu, Z., Tembo, G., Adamba, C. & Mvula, P. (2023) 'An empirically driven theory of poverty reduction', CEDIL Research Project Paper 4. Centre of Excellence for Development Impact and Learning (CEDIL), London and Oxford. Available at: https://doi.org/10.51744/CRPP4

Hydén, G. (1983) *No Shortcuts to Progress*. Orlando, Florida: University of California Press and Heinemann.

Hydén, G. (1980) *Beyond Ujamaa in Tanzania: Underdevelopment and an Uncaptured Peasantry*. Orlando, Florida: University of California Press and Heinemann.

Merton, R. K. (1949) *Social Theory and Social Structure*. New York: Free Press.

Oliver, S., Roche, C., Stewart, R., Bangpan, M., Dickson, K., Pells, K., Cartwright, N., Hargreaves, J., and Gough, D. (2018) 'Stakeholder Engagement for Development Impact Evaluation and Evidence Synthesis'. *CEDIL Inception Paper 3*. Oxford: CEDIL. Available at: https://doi.org/10.51744/CIP3

Oliver, S., Langer, L., Nduku, P., Umayam, H., Conroy, K., Maugham, C., Bradley, T., Bangpan, M., Kneale, D., Roche, C. (2021) 'Engaging Stakeholders with Evidence and Uncertainty: Developing a toolkit'. *CEDIL Methods Working Paper 4*. Oxford: CEDIL. Available at: https://doi.org/10.51744/CMWP4

Pawson, R. (2006) *Evidence-Based Policy : A Realist Perspective.* London: Sage

Pawson, R. & Tilley, N. (1997). Realistic Evaluation. London: Sage.

Vogel, I. (2012) 'Review of the use of "Theory of Change" in international development'. London: DFID. Available at: "www.theoryofchange.org/pdf/DFID_ToC_Review_VogelV7.pdf

Weiss, C.H. (1997), Theory-based evaluation: Past, present, and future. New Directions for Evaluation, 1997: 41-55. Available at: https://doi.org/10.1002/ev.1086

White, H. (2005) *Maintaining Momentum to 2015? An impact evaluation of interventions to improve maternal and child health and nutrition in Bangladesh*. Washington DC, OED and World Bank.

White, H. (2010) 'Theory-based impact evaluation: principles and practice'". *Journal of Development Effectiveness*, 1(3), pp. 271–284. Available at: https://doi.org/10.1080/19439340903114628

White, H. and Masset E. (2006) 'Assessing interventions to improve child nutrition: a theory-based impact evaluation of the Bangladesh Integrated Nutrition Project'. *Journal of International Development* 19, pp. 627–652. Available at: https://doi.org/10.1002/jid.1344

White, H. and Masset, E. (2018) The rise of impact evaluations and challenges which CEDIL is to address, *Journal of Development Effectiveness*, 10:4, 393-399. Available at: https://doi.org/10.1080/19439342.2018.1539387

Annex 1: Overview of CEDIL projects using middle-level theory

Project name	MLT element	Developing theory	Method	Use
An empirically driven theory of poverty reduction	Apply machine-learning algorithms to identify the variables that predict large gains in consumption among cash transfer recipients. By looking across four different settings, we can use the empirical results to construct a mid-level theory of graduation from ultra-poverty	Bottom up	Quantitative	Theory development
Climate aid	Aggregate across intervention-level ToCs to develop a middle-level theory	Bottom up	Qualitative	Develop theory
Demand for contraception among adolescents	An integrative, mixed-methods review to build a mid-range theory of adolescent contraceptive demand generation	Bottom up	Formalised qualitative	Theory development and testing
Developing theory and methods for evaluating government training on citizen engagement	Use process-tracing case studies to develop a middle-level theory for governance interventions focused primarily on capacity building of community leaders by specifying how, where, and when government training is effective at improving government effectiveness, inclusion, and legitimacy	Bottom up	Qualitative	Develop theory
Gender and social outcomes of WASH interventions	Advance the use of mixed methods with logic models to hypothesise causal relationships among intervention components. The logic model will be built iteratively, informed by theory, and refined by stakeholders	Iterative	Qualitative	Theory development and testing
Impact evaluation of the SHARPE project in Ethiopia	Create a middle-level theory by using randomised encouragements to create experimental variation in participation to observe and measure how well benefits carry through the causal chain across different regional contexts	Bottom up	Quantitative	Develop theory

CEDIL Lessons Learned Paper 1: The use of middle-level theory in CEDIL-funded research studies

Integrating data and theory for better inference	Develop an accessible approach to building and communicating middle-level theory using data by using simple graphical methods (DAGs) to build 'thick descriptions' of the context in which interventions will interact	Bottom up	Qualitative	Develop theory
Involving men and boys in family planning	Test and adapt the logic model to identify which family planning interventions for men and boys work, for whom, and under what circumstance by using a CCA approach to testing of causal pathways and identifying system- and process-level barriers and facilitators to effective intervention	Top down	Formalised qualitative	Testing theory
Language transitioning research synthesis	Build a middle-level theory based on the psycholinguistic underpinnings of reading and sociolinguistic contexts of learning that may explain why certain language of instruction policies are likely to be more effective than others	Top down	Qualitative	Testing interventions
Machine-learning methods to uncover mechanisms underlying the impacts of two long-term evaluations of youth skills training programmes	Apply machine-learning regression tree methods to inform the generalisability of the observed impacts and also expand on machine-learning methods to conduct causal mediation analysis, which will contribute to the formulation of middle-level theories	Bottom up	Quantitative	Develop theory
Predicting optimal policies for new contexts using existing studies	Represent middle-level theories as structural economic models of behaviour fitted to pre-existing experiments and descriptive data. The economic models generate predictions for various counterfactual scenarios and can thus inform policy recommendations	Top down	Quantitative	Predicting impact
Scaling social accountability for health	Develop a middle-level theoretical framework to evaluate how and under what contextual conditions development outcomes are improved where scaled-up social accountability is incorporated as a sustainable part of development policymaking and implementation. This looks beyond the results of individual projects, in terms of information provision, to why and how they	Bottom up	Qualitative	Develop theory

	are integrated within the political economy of policymaking processes in health and to what effects			
Structural estimation of spatial spillover effects of cash transfers	Develop and apply a middle-level theory of spillover effect in the presence of partial economic integration of local markets and so draw policy conclusions from the findings that are relevant in broader contexts than the specific intervention evaluated	Iterative	Quantitative	Theory development and predicting impact
Syrian business development and regional trade in a humanitarian setting	Combine quantitative data and qualitative data to guide the development of middle-level theory on buyer–seller trade across countries in a conflict/post-conflict setting	Top down	Quantitative	Develop theory
The art and science of using evidence	Use rigorous research synthesis – informed by a theory-based approach – on what works and in what contexts in relation to six key types of efforts to increase the use of evidence in policymaking in different contexts	Top down	Qualitative	Test interventions
Transferability of education mid-range theories	Meta-analysis using a pre-specified set of coded study features with moderator variable analysis to answer questions about the transferability of evidence	Top down	Quantitative	Predicting impact
Understanding factors that influence TaRL's effectiveness and generalisability	Combines an applied theoretical model with quantitative results from a Bayesian meta-analysis to understand generalisability. The conclusions of the mid-level theory to enter the formal model via the priors on the meta-regression coefficients as well as context and programme component interaction terms	Top down	Quantitative	Predicting impact
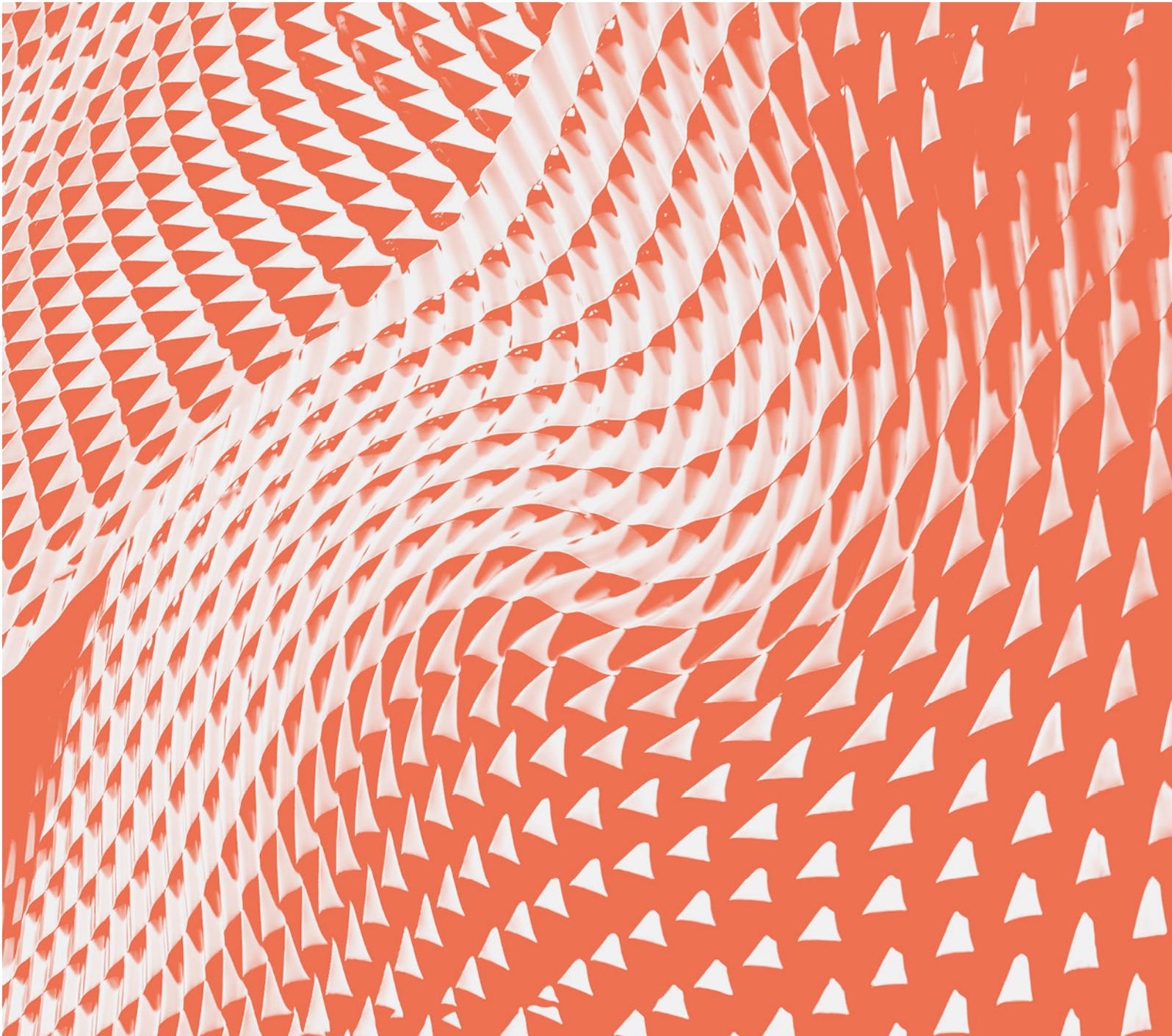
About CEDIL

The Centre of Excellence for Development Impact and Learning (CEDIL) is an academic consortium supported by the UK Government through UKaid. The mission of the centre is to test innovative methodologies in evaluation and evidence synthesis and to promote evidence-informed development. CEDIL-supported projects fall into three programmes of work: evaluating complex interventions, enhancing evidence transferability, and increasing evidence use.

For more information on CEDIL, contact us at **cedil@opml.co.uk** or visit our website **www.cedilprogramme.org**