

KEY POINTS

- Whether a policy that worked in one context will also work in another depends on the interaction of the policy's mechanism with features of the context. This is its "external validity".
- Existing methods are helpful for understanding a policy's external validity, but are limited to examining a small number of dimensions of context.
- Mechanism mapping is a five-step procedure to help policymakers identify potential mechanism-context interactions, and come up with suitable adaptations.

PEGNet

Policy Brief

External Validity and Policy Adaptation: From Impact Evaluation to Policy Design

Martin J. Williams

With the recent boom in impact evaluations around the world, policymakers in many sectors now have at their disposal an overwhelming amount of evidence about "what works" - or at least what worked in a particular context. Yet as impact evaluations have multiplied, it has become apparent that "the same" policy can have very different effects in different populations (Vivalt 2019). Similarly, policies shown to be effective in small trials have not always been as effective when implemented at scale, even in the same country (Bold et al 2018).

This is the problem of the external validity. The limited external validity of impact evaluation evidence poses challenges for policymakers: how can one know if a policy will have the same effect in this implementation context as it did elsewhere? And to what extent should policymakers copy the design of policies that have worked elsewhere, rather than use local information to try to adapt them to fit the local context?¹

In a recent paper (Williams 2019), I propose a simple and flexible framework for thinking about these questions, and about external validity more broadly. A policy can have a different impact in a new context than it had in a previous context if part of a policy's theory of change (a.k.a. its mechanism) interacts with a difference in contexts. A policy's theory of change is a mapping of its intended mechanism spanning inputs to activities, outputs, intermediate outcomes, and final outcomes.

Whether this mechanism works as intended depends at each step on the validity of a set of contextual assumptions. While these assumptions may have been true of the context in which a policy had previously been shown to work, whether the policy will have the same effects in a new context depends on whether these same contextual assumptions hold. Since context can include a wide range of factors - location, target group, implementing organization, scale, time period, the existence of related policy interventions, etc. - and the theory of change includes

factors related to implementation as well as impact, this simple framework can be used to analyze a wide range of factors affecting policy impact.

This policy brief summarizes a range of existing approaches to external validity in economics and other social sciences, and how they relate to this perspective. It then summarizes a new approach, called mechanism mapping, to analyzing whether a policy that worked in another context will work in a new context.

Existing Approaches to External Validity

The existing academic literature on external validity provides increasingly insightful answers to questions of the generalizability of impact evaluations or bodies of evidence - whether evaluation results from a specific context will hold in unspecified other contexts. However, it provides more limited insight into concerns about applicability of evidence - whether evaluation results from various other contexts will hold in the specific context in which a policymaker is working.

One empirically driven response to the variability of policy impacts across contexts is to aggregate numerous studies of the same policy. As the policy is tried and evaluated in more contexts, it may become possible to aggregate these results further into an average treatment effect across studies, through a systematic review or a meta-analysis. But this estimate is of an average treatment effect in the average context in which the policy has been evaluated, which can differ from the policy's effect in a specific new context in two ways. First, the populations in which the policy has previously been tried and/or evaluated may differ systematically from the new context in important ways. For many social policy interventions, for example, there exist numerous studies from OECD countries but little or no evidence in developing countries, and Allcott (2015) has shown that policy experiments are often conducted first in the most favorable locations, leading to a site selection bias effect. Second, there can be significant heterogeneity in policy impact



KIEL INSTITUTE FOR THE WORLD ECONOMY
Kiellinie 66 | 24105 Kiel

E pegnet@ifw-kiel.de
www.pegnet.ifw-kiel.de

@pegnet_

¹ This policy brief is based on and adapted from a working paper (Williams 2017a), policy memo (Williams 2017b), and World Bank Research Observer article (Williams 2019).

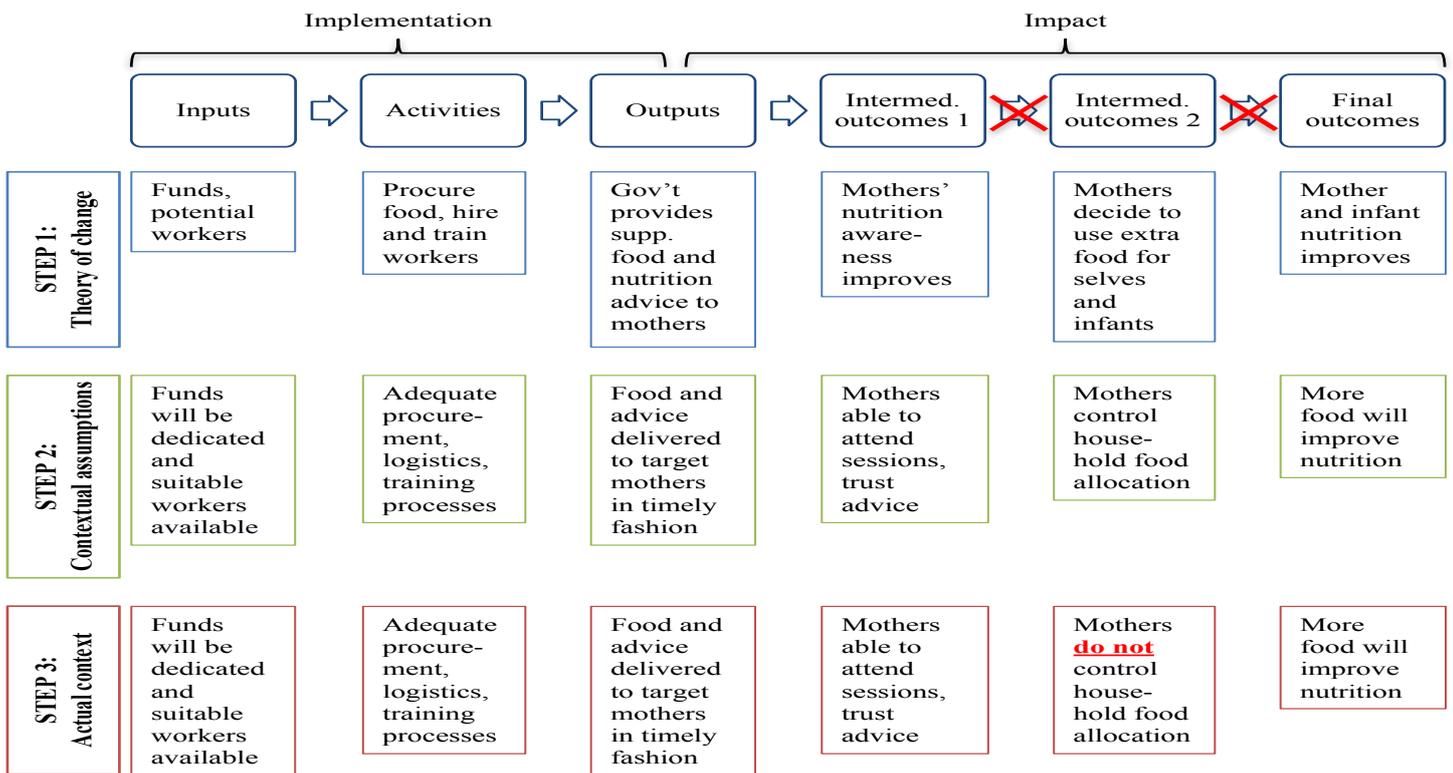


Figure 1: Mechanism Map for Bangladesh Integrated Nutrition Programme (BINP)

across contexts, so that a policy that has a positive effect on average could have a negative effect in some contexts.

The main empirical approach to dealing with heterogeneous effects is to employ sub-group analysis and/or interaction effects, which allow the researcher to see how average treatment effects vary across important variables. But such analysis is inherently limited in the number of variables along which it can disaggregate results. Inevitably, there will be some contextual variables that influence a policy's effectiveness that are either difficult to measure or that evaluators might not think to measure ex ante and are thus unobserved. While systematic reviews can be a useful starting point for policymakers, naively adopting a policy that has a positive "headline" average treatment effect in a systematic review is likely to backfire in many contexts.

A second approach focuses on making out-of-sample extrapolations through structural modelling (Deaton and Cartwright 2016) or other empirical methods that can, in some circumstances, be used to extrapolate results from one study to other populations, by exploiting specific forms of selection and non-compliance within RCTs or by adjusting estimated impacts based on heterogeneity over observed covariates (Angrist and Fernandez-Val 2010, Gechter 2016, Kowalski 2016, Andrews and Oster 2018). These methods help researchers and policymakers further improve the informativeness of the existing evidence about the predicted impact of the intervention in a new context, but are also inherently limited in the number of variables and types of scenarios across which they can extrapolate. This is certainly informative for policymakers in specific contexts, and is an improvement over simply having the results of an impact evaluation from another context without such extrapolation, but still falls short of taking into account all the potential mechanism-context interactions with which policymakers must concern themselves.

Similarly, the design of policy experiments can vary aspects of the policy that are important for understanding external validity, such as whether it is implemented by an NGO or government (Cameron and Shah 2017, Angrist 2017, Bold et al 2018). Again, the limitation is that trials can only feasibly vary one or two dimensions of a policy without losing statistical power, while the number of dimensions of policy and context that could matter - combined with their interactions - is effectively infinite.

This range of existing methods provides a powerful set of tools to analyze external validity, and together they can help policymakers select policies that are more likely to be effective. But they also have important limitations – in particular the inability to deal with the high dimensionality of policies and contexts, and the important role that context-specific "unobservables" play in determining policy effectiveness. Mechanism mapping is designed to address these limitations and thus to be used as a complement to these existing methods.

Mechanism Mapping in Five Steps

The process of mechanism mapping can be broken down into five steps. I illustrate the approach using Cartwright and Hardie's (2014) example of the Bangladesh Integrated Nutrition Programme (BINP), a mainly World Bank-funded project in the 1990's. The design of BINP was copied exactly from the World Bank's highly successful Tamil Nadu Integrated Nutrition Programme (TINP) – a clear example of evidence-based policy – yet BINP had little impact on its key outcomes. The mechanism map below will make clear how the same policy could be so effective in Tamil Nadu but ineffective in Bangladesh.

Step 1: Map out the policy's intended Theory of Change (i.e. mechanism)

The first step of mechanism mapping is to lay out a policy's theory of change, or mechanism. This can be thought of as a causal chain leading from a policy's initial inputs to its intended final outcomes, via activities, outputs, and intermediate outcomes. For clarity, this article explains mechanism mapping using a simple, linear theory of change, but the mechanism mapping process can be used with whatever style of theory of change the analyst prefers.

The intended final outcome of BINP was to improve mother and infant nutrition. To do so, government was to provide two main outputs: nutritional advice delivered to pregnant and nursing mothers, and the distribution of supplementary food to mothers to take home. These outputs would lead to the final outcome via two sets of intermediate outcomes: first, mothers' nutritional awareness would improve, alongside their receipt of the supplemental food; and second, mothers would then decide to use the supplemental food for themselves and their infants (as opposed to giving it to other family members, i.e. program "leakage"). In order to produce these outputs, the government required inputs of adequate financial resources to purchase the food and pay personnel, as well as a logistical system and potential pool of extension workers to deliver the food and nutritional advice. Key activities for transforming inputs into outputs could include procuring the food, hiring and training workers, and conducting outreach to eligible mothers.

Step 2: Map the contextual assumptions that must hold for the Theory of Change to work

The contextual assumptions required for this theory of change to work are listed in the second row of Figure 1. Sound implementation requires that government: dedicate adequate financial and human resources to the project; procure and distribute food and hire workers effectively, including quality assurance as well as prevention of excessive corruption, and train workers adequately; and deliver these outputs to a pool of eligible mothers predictably and in a timely fashion. Impact then requires that mothers are able to attend the sessions and trust the advice they are being given; that mothers actually control household food allocation; and that the supplementary food, if consumed, will actually lead to the desired improvement in nutrition. In the Tamil Nadu context, these assumptions were presumably valid - hence the impact evaluation finding that TINP significantly improved mother and infant nutrition (World Bank 2005b).

Step 3: Map the actual characteristics of the context and compare them to the assumptions

The third row of Figure 1 contrasts these contextual assumptions to the actual contextual characteristics of the new context, in this case rural Bangladesh. The key contextual assumption that did not hold in Bangladesh was that mothers controlled household food allocation and would thus be able to act on their improved nutritional awareness: whereas mothers were typically responsible for shopping and household food allocation decisions in rural Tamil Nadu, in rural Bangladesh men usually conducted the shopping and their mothers (the mothers-in-law of the pregnant or nursing women) controlled household food allocations (White 2005; Cartwright and Hardie 2014).

This broke the link between Intermediate Outcome 1 and Intermediate Outcome 2: while BINP succeeded in distributing food and nutritional advice to the mothers, and mothers' nutritional awareness did actually improve as a result, the program failed to improve mother and infant nutrition because most of the supplementary food went to other family members. Since Intermediate Outcome 2 was not achieved, neither was the Final Outcome. If the designers of BINP had carried out a mechanism mapping when transporting the successful TINP program to Bangladesh, perhaps they would have uncovered this crucial but implicit assumption.

Mechanism mapping can also be adapted to policies that are intended to lead to multiple final outcomes (e.g. a cash transfer that is intended to increase consumption and improve child school attendance) simply by creating multiple mechanism maps, one for each outcome. The theory of change may be the same for each outcome or may differ slightly in emphasizing the aspects of the mechanism that are more salient, but the key contextual assumptions and characteristics are likely to be different. The same procedure can also be used to assess the likelihood of negative outcomes or side effects of the policy, by placing these undesirable outcomes as the final outcome of the policy and assessing whether the policy mechanism and contextual characteristics and assumptions are likely to lead to them.

Empirical evidence has an important role to play in mechanism mapping. Most obviously, the contextual characteristics in the crucial bottom row are questions to which empirical answers - or at least suggestive evidence - may well exist. Bates and Glennerster (2017) gives some excellent examples of the use of evidence to identify contextual differences prior to transporting a program.

Step 4: Adapt the policy to eliminate these mismatches between assumptions and characteristics

Since mechanism mapping as a diagnostic tool focuses on the interaction between a policy's theory of change and differences in context, the diagnosis of whether a policy is likely to be as effective in a new context as it was elsewhere inherently involves highlighting the aspects of the policy that should be targeted for adaptation. In the case of BINP, for example, Figure 1 makes it obvious that the key aspect where adaptation was necessary was the nutritional advice component, and specifically the individuals to whom this was targeted.

An obvious way to adapt the policy would be to extend the nutritional advice component to include the key decision makers about household food allocations besides mothers - their husbands and their mothers-in-law. This may also require changes to other parts of the program, since these new target populations may have to be reached in different ways, for instance through home visits. This in turn may imply other changes to the policy's theory of change, both in terms of effective implementation (greater resources needed, additional logistical and transportation issues) and also for these advice sessions to have the intended impact (e.g. home visits may raise different cultural or trust issues).

Of course, many plausible adaptations could be made. As the following step discusses, mechanism mapping can be used to compare the feasibility and likely effectiveness of these options in a systematic fashion.

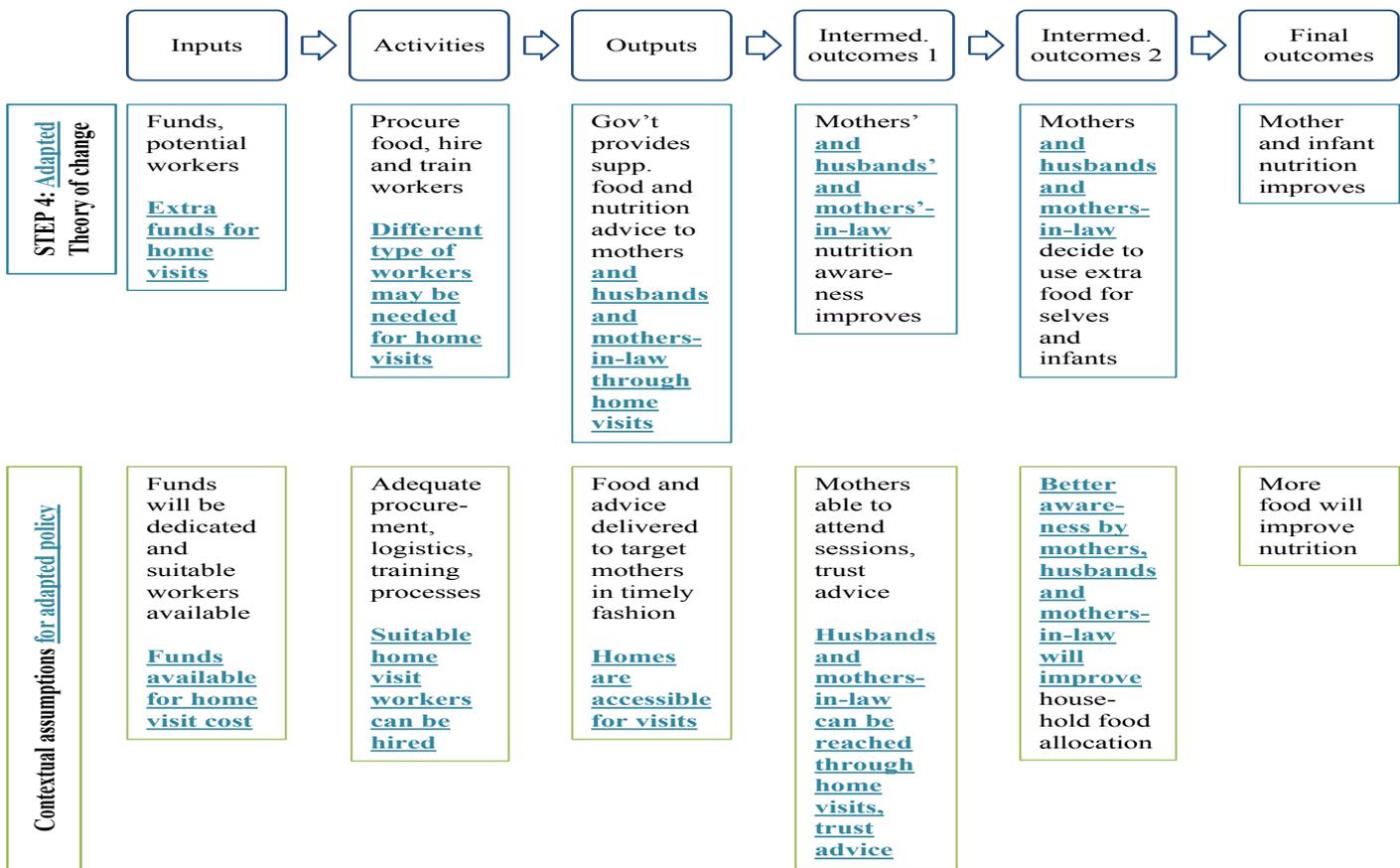


Figure 2: Adapted Theory of Change and Assumptions for BINP

The other benefit of using mechanism mapping to suggest adaptations is that it makes it clear what aspects of the policy do not need to be adapted. For instance, Figure 1 makes clear that the other steps in BINP’s theory of change fit well with the contextual assumptions and previous context in which the program had been evaluated, suggesting that there is little need for adaptation in these respects (except as necessitated by the adaptations in Figure 2). The design of the resulting adapted policy is thus informed both by evaluation evidence from other contexts - through the aspects of the original policy that were maintained in the new context - as well as by local, context-specific knowledge - through the aspects that were adapted.

Step 5: Repeat steps 1-4 for the adapted policy, iterating in more detail until satisfied that all major policy design decisions fit the local context

While the adaptation process can help mitigate the contextual differences – and possibly even improve on the original program – these adaptations also require new contextual assumptions to be effective, and these contextual assumptions must also be evaluated. Having proposed adaptations, the analyst should therefore repeat the mechanism mapping process in steps 1-4 for the adapted policy – or, in the case of multiple alternative adaptations, for the multiple versions of it. Although this policy memo does not illustrate this iteration process for the sake of brevity, iterating in this way is crucial to help the analyst to assess the feasibility and likely impact of each adaptation. Where the contextual assumptions and characteristics do appear to match, the iteration process can be used to drill down into more detail to be more precise about exactly what adaptations are (or are not) needed.

Conclusion

Mechanism mapping is a simple and flexible tool to help policymakers identify external validity failures and design adaptations to address them. While evidence-based policymaking might use a successful impact evaluation from another context or a systematic review as a starting point for policy design, mechanism mapping can help policymakers make the adaptations necessary for the policy fit in their specific context.

While mechanism mapping is intended primarily as a tool for policymakers, mechanism mapping is also of potential value to evaluators in two ways. First, it can be useful in the retrospective evaluation of policies by helping evaluators to show clearly the intended and actual mechanism(s) through which a policy had its impact (or non-impact). Second, prospective mechanism mapping can also help evaluators design trials to ensure that they collect the data necessary to assess each of the contextual assumptions ex post, along with potential undesirable outcomes and the alternative mechanisms that might bring them about.

Finally, mechanism mapping should be understood as a tool to help policymakers structure their judgment about policy transportation and adaptation, not a scientific procedure for determining whether or not a policy will work. It relies on policymakers’ judgment, but seeks to structure and improve it in the pursuit of better use of evidence in policymaking.

Author

Martin J. Williams
Blavatnik School of Government
University of Oxford
martin.williams@bsg.ox.ac.uk

PEGNet Policy Briefs

provide information, analysis and key policy recommendations on important development and humanitarian topics. The views presented are those of the authors and do not necessarily reflect the views of PEGNet. In case of questions or comments, please directly contact the author.

References

- Allcott, H. 2015. "Site Selection Bias in Program Evaluation." *Quarterly Journal of Economics* 13 (3): 1117–65.
- Andrews, I., and Oster, E. 2018. "Weighting for External Validity." NBER Working Paper 23826, National Bureau of Economic Research, Cambridge, MA.
- Angrist, N. 2017. "An Application of the Jump from Internal to External Validity: Transporting an HIV Prevention Intervention from Kenya to Botswana, and Testing Two Scale Models Across Key Parameters of Heterogeneity." Working Paper, Blavatnik School of Government, Oxford University, Oxford, UK.
- Angrist, J., and Fernandez-Val, I. 2010. "Extrapolating: External Validity and Overidentification in the LATE Framework." NBER Working Paper 16566, National Bureau of Economic Research, Cambridge, MA.
- Bates, M. A., and Glennerster, R. 2017. "The Generalizability Puzzle." *Stanford Social Innovation Review*, Summer.
- Bold, T., Kimenyi, M., Mwabu, G., Ng'ang'a, A., and Sandefur, J. 2018. "Experimental Evidence on Scaling Up Education Reforms in Kenya." *Journal of Public Economics* 168, 1–20.
- Cameron, L., and Shah, M. 2017. "Scaling Up Sanitation: Evidence from an RCT in Indonesia." IZA Discussion Paper 10619, Institute of Labor Economics, Bonn, Germany.
- Cartwright, N., and Jeremy H.. 2014. *Evidence-Based Policy: A Practical Guide to Doing It Better*. Oxford: Oxford University Press.
- Gechter, M. 2016. "Generalizing the Results from Social Experiments: Theory and Evidence from Mexico and India."
- Kowalski, A. E. 2016. *How to Examine External Validity Within an Experiment*.
- Vivalt, E. 2019. "How Much Can We Generalize from Impact Evaluations?" Mimeo, February.
- White, H. 2005. "Comment on Contributions Regarding the Impact of the Bangladesh Integrated Nutrition Project." *Health Policy and Planning* 20(6): 408–11.
- Williams, M. 2017a. "External Validity and Policy Adaptation: From Impact Evaluation to Policy Design." BSG Working Paper 2017-019, July. Available at: <<https://www.bsg.ox.ac.uk/sites/default/files/2019-06/BSG-WP-2017-019.pdf>>
- Williams, M. 2017b. "External Validity and Policy Adaptation: A Five-Step Guide to Mechanism Mapping." BSG Policy Memo 2017-07, July. Available at: <<https://www.bsg.ox.ac.uk/sites/default/files/2018-06/2017-07%20Policy%20memo%20External%20validity%20and%20policy%20adaptation%20%20Williams%20.pdf>>
- Williams, M. 2019. "External Validity and Policy Adaptation: From Impact Evaluation to Policy Design." Forthcoming, World Bank Research Observer.
- World Bank. 2005a. "The Bangladesh Integrated Nutrition Project: Effectiveness and Lessons." *Bangladesh Development Series Paper No. 8*, December.
- World Bank. 2005b. "Project Performance Assessment Report: Bangladesh Integrated Nutrition Project." Report No. 32563, 13 June.

