

# Scaling policy ideas in developing countries

**John A. List**  
**U. Chicago, ANU, and NBER**

Teaser: To identify effective policies that can scale, a third option should be added to traditional A/B tests, that accounts for the realities of a programme implemented at scale. By flipping the traditional research and policy-development model, researchers can generate policy-based evidence to help policymakers scale the best policies.

Policymakers in developing countries are increasingly relying on experimental methods to tackle social challenges. A crucial obstacle in this endeavor is the difficulty of scaling successful interventions from a small, controlled setting to larger, more diverse populations. The urgency of scaling policies affects us daily, whether it's in safeguarding the health and safety of communities, enhancing the sustainability of development initiatives, or improving educational opportunities for future generations.

While its importance is undeniable, the scaling process is fraught with challenges, from the inception of an idea to well after policy implementation. One critical consequence of these challenges is the "voltage drop", where the benefit-cost profile of an idea often depreciates significantly when moving from a small scale to a larger, more diverse context (List 2022). For example, elegant solutions to combat climate change may work in controlled environments but fall apart when scaled. Similarly, educational reforms may show promising results in pilot programmes but see diminished effects when implemented nationwide. In my research, I have identified that some ideas are predictably scalable, while others are predictably unscalable, with their scalability determined by factors visible through a scientific lens.

Today, many policymakers, in a range of circumstances and geographies across the world, prioritise developing "evidence-based policies". In many circles, this involves advancing public policies, programmes, and practices grounded in empirical evidence, often without considering how relevant that evidence is to the specific contexts of what, for whom, where, and how they aim to scale. To address voltage drops, I argue that we must optimally generate policy-based evidence before decision-makers commit to scaling and continue to do so throughout the policy's lifecycle.

Optimality requires recognising two key features: **identifying potential scaling threats** and **determining the information needed to provide scaling confidence**. For policymakers, this necessitates flipping traditional research and policy-development models. To achieve this, original experimental designs and prototyping must address potential stumbling blocks from the outset, engaging in backward induction. In the traditional model of "learning by doing," scholars often only consider scaling ideas and interventions after they demonstrate success in efficacy tests, commonly referred to as "A/B tests" in the experimental community.

Yet, it is often economically and scientifically efficient to not only conduct the efficacy test but also relevant tests of scale and mechanisms within the original discovery process. In its simplest form, this means introducing Option C Thinking to the standard experimental approach. This approach effectively augments the typical A/B test by including a scalable version of the studied program alongside a "best case" version (List 2024). Leveraging Option C Thinking in our initial designs moves us one step closer to flipping the traditional social science research model from efficacy trials to an approach that produces the type of policy-based evidence that the science of scaling demands.

### **Before scaling a policy, use ‘Option C Thinking’ to provide *Policy-Based Evidence***

A hallmark of public policy decision-making is a comparison of the benefits and costs associated with proposed programmes or regulations. In this spirit, as knowledge creators interested in scaling, we must answer the following question: after a programme has been claimed to pass a benefit-cost test in an initial study, what is the probability that it will pass a benefit-cost test in the target setting of interest?

A crucial aspect of answering this question involves understanding how information from the experimental setting translates to the scaled setting. The challenge lies in the fact that social science experiments involve humans, making it nearly impossible to develop behavioural laws akin to those in the natural sciences. Humans are creatures of habit, until they are not. We respond predictably to stimuli, until an unnoticed factor changes and triggers an apparently irrational reaction. Consequently, behavioural principles observed in one environment do not always apply broadly. This inherent complexity makes humans maddening experimental subjects and renders the social sciences the "harder sciences."

## **Five key threats to scaling policies**

Leveraging a set of economic models and the voluminous empirical literature, in List (2022), I outline five threats that can cause voltage drops and prevent an idea from having its promised impact when scaled. I denote these as the 5 Vital Signs, and they can be summarised as follows:

### *Vital Sign #1: False positives*

False positives occur when an idea appears successful in an experimental setting, but the idea never had true potential (or "voltage") in the first place.

### *Vital Sign #2: Representativeness of the population*

In this case, the voltage effect arises when the decision-maker assumes that the small subset of people who are affected by the policy in the Petri dish are more representative of the general population than they are at scale.

### *Vital Sign #3: Spillovers*

The third vital sign is an understanding that the implementation of the policy can have unintended consequences, or spillovers, that work against (or in favour of) the desired outcome.

### *Vital Sign #4: Supply side*

The fourth vital sign is the “supply side” of scaling—if a policy has diseconomies of scale as it expands it becomes costlier to sustain.

### *Vital Sign #5: Representativeness of the situation*

Our last key component of scaling is representativeness of the situation. This is a broad and rich category that includes any situational feature that impacts the effects of the programme at scale, including mediators and moderators.

Option C Thinking is considering each of the Vital Signs and exploring the ones that pose the greatest threat to your idea or policy to scale.

## **Introducing Option C Thinking to A/B Testing**

A general lesson from my work is that recognising which of the 5 Vitals represent the biggest threats to scaling and thoroughly exploring them is crucial for gaining scaling confidence. For instance, let's say strong results are found from a home visiting program that alters parental beliefs (List et al. 2021). Indeed, we find that such treatments have significant impacts on early childhood investment and child outcomes. The next step involves identifying potential key factors that could influence the programme's scalability. This determination can be made using local knowledge, previous results, theory, or even Generative AI. By identifying and addressing these critical factors early on, researchers and policymakers can better prepare for the challenges of scaling and ensure that interventions retain their effectiveness when expanded to larger, more diverse populations.

Say in this process you determine that the biggest scaling threats are i) the relevant mediation path and ii) the quality of the home visitor. In terms of the mediation path, List et al. (2021), had a complex causal pathway. For this idea to be scalable, it is important to recognise that their approach relies on: i) the treatment changing parental beliefs, ii) parents, whose beliefs change, having adequate resources to invest in child, iii) those investing parents understanding how to invest in the child, iv) the parents investing accordingly, and v) those investments moving the child outcome of interest within the experimental time-frame.

If this entire chain holds in the target setting of interest with the fluidity that it held in the experimental setting, then the programme passes the mediation path test.

This example highlights a common approach in development circles. A key consideration for scaling ideas in low- and middle-income countries is that key features of the situation often change across different regions. Behavioural biases such as present bias, where individuals favour immediate rewards over future benefits, or a lack of trust in institutions, can further complicate scaling efforts by altering how populations respond to interventions. While dividing a complex problem into several distinct smaller parts has its merits, it's crucial to understand the complete set of mechanisms—including behavioural influences—before generalising insights from a part of the puzzle to the whole. Without a clear understanding of how the entire puzzle fits together (including mediators and moderators),

one should be cautious about generalising and scaling ideas. The overarching message is that creating ideas with simple mediation paths is vital. Greater trust should be given to policies that are well understood theoretically, address context-specific behavioural biases, and have straightforward mediation paths.

Now, consider the second scaling threat—quality of the home visitor. If upon backward-induction we realise that at scale in millions of households and thousands of different settings, our programme wouldn't have its ideal applicant pool of home visitors to choose from, we must explore this facet of the scaling problem. To do so, we should design our initial experiment to test if our programme works with home visitors who have varying abilities. In practice, this means using home visitors who would typically be part of the scaled programme in the original test. This choice naturally provides the A/B efficacy test with several stellar home visitors, but also incorporates Option C Thinking, ensuring that the situation is representative, at least in terms of home visitor quality. While it might sound counterintuitive, or even foolish, to not seek out and hire only the best talent in the early stages, this approach provides the insights necessary for scaling right from the outset.

This approach provides a home visitor pool that allows for a heterogeneity test on this factor. Of course, this example focuses on home visitors to understand the programme, but beyond home visitors, a wealth of other considerations could be examined, as discussed above with the 5 Vital Signs.

### **Scaling policy ideas in low- and middle-income countries**

A novel and promising approach to applying Option C thinking in developing contexts is incorporating the perspectives of local stakeholders during the design phase of interventions. This method, as demonstrated by Dal Bó et al. (2021) in Paraguay, engages those closest to the ground to provide input on which treatments are most effective in their specific contexts. By leveraging local insights, researchers can tailor interventions to fit local realities, thereby increasing their chances of success at scale. This participatory methodology serves as a valuable tool for development economists, enabling them to address context-specific challenges early and enhance the scalability of interventions in diverse and resource-constrained settings.

The larger point here is that there are key cases where the researcher should backward induct and explore crucial elements and constraints that the idea will face at scale. In its most basic form, I refer to this approach as Option C Thinking: creating evidence that provides greater scaling confidence in the original design alongside the efficacy test.

Dal Bó, Ernesto, et al. "Information technology and government decentralization: Experimental evidence from paraguay." *Econometrica* 89.2 (2021): 677-701.

List, J.A., "The Voltage Effect," Crown Currency Publishing, (2022).

List, J.A., "Optimally Generate Policy-Based Evidence Before Scaling," *Nature*, 626: 491–499 (2024).

List, J.A., Pernaudet, J. & Suskind, D.L. Shifting parental beliefs about child development to foster parental investments and improve school readiness outcomes. *Nat Commun* **12**, 5765 (2021).