The Science of Scaling FAQ

What is the science of scaling?

The science of scaling is, in essence, the science of using science. It is a research field dedicated to examining how experimental insights can effectively be scaled and to advancing the idea that experimental research must be applied in a careful, systematic way in order to reap the benefits found within.

The science of scaling represents the next frontier in evidence-based policy making. As policy makers increasingly turn to insights gained from the experimental method as a means of informing public policies, whether—and to what extent—those insights are likely to scale to the level of the broader public is often based on blind faith. And in reality, many interventions that deliver promising results in research settings fail to deliver the same positive effects when scaled up to a broader population.

Understanding why interventions that work well in initial research studies then fail to scale up is the critical question at the heart of the science of scaling.

Most research studies, even rigorous randomized controlled trials (RCTs), only reveal what an intervention does for a particular population and in a particular situation. A promising RCT does not mean that its results can be applied to other children or families or in different communities. The science of scaling helps reveal what factors determine whether an intervention might deliver the same level of impact in a different place, in a different situation, and with a different population.

What is the scale-up effect?

The “scale-up effect” refers to the change in magnitude of a treatment effect when an intervention moves from the research setting to population-wide implementation. The term was coined by Omar Al-Ubaydli, John List, Claire Mackevicius, Min Sok Lee, and Dana Suskind to refer to net changes in the benefit/cost profile resulting from changes in scale, as opposed to the more commonly discussed “voltage drop” phenomena, which considers only changes in benefits (but not costs) brought about by the scaling of a program. The scale-up effect can be negative or positive and can arise from the benefit or cost side of a program.

What is the economic model of the science of scaling?

In their 2017 paper, “What Can We Learn from Experiments? Understanding the Threats to the Scalability of Experimental Results,” Al-Ubaydli, List, and Suskind present an economic model, or theoretical framework, for understanding the science of scaling and what causes the scale-up effect.

The model identifies and considers the incentives of three main players, the government, researchers, and the populace, as they relate to a new proposed intervention or program. (Scientists study the program, while the government follows the research findings with an eye toward implementing the program, and the populace participates in the program.) The program, as defined in this model, leads to a direct per capita treatment effect TT, and it has a per capita cost CC. The per-capita net treatment effect, which measures the program's impact net of costs, is: $\tau\tau = TT - CC$

Having established the per-capita net treatment effect, the authors then use economic modeling to explore the scale-up effect, that is, changes in the magnitude of $\tau\tau$ when the program moves from the research setting to population-wide implementation.

This allows them to identify the specific sources of the scale-up effect, which they sort into four key categories:

1. Errors in Statistical Inference - Researchers and policymakers do not always make the correct inference from available data. The initial, promising research results may have been interpreted incorrectly, meaning there is not sufficient evidence to support scaling.
2. Properties of the population - Individuals studied in the research setting may not be representative of the population at-large. A non-representative participant pool can be caused by several factors, including the fact that in the scientific marketplace, researcher incentives dictate a subject pool choice that is more likely to find larger treatment effects than would a random sample.
3. Properties of the situation - The specifics of the program and the ways in which it is delivered/received in the research setting may not be representative of the broader real-world context.
4. Spillover and general equilibrium effects - In some interventions, treating people creates a spillover effect on others (others receiving the treatment, individuals in the control group, or even people who are not participating in the experiment at all). These effects can be negative or positive and can pose a threat to scalability.

How does this economic model of the science of scaling differ from implementation science?

The fields of implementation science and the science of scaling are complementary fields that both contribute toward tackling the challenge of scaling.

While the implementation science literature is deep, it typically revolves around the "voltage effect"—the conjecture that treatment effect sizes observed in research studies diminish substantially when the program is rolled out at larger scale (Kilbourne et al., 2007; Weiss et al., 2014; Supplee and Meyer, 2015; Supplee and Metz, 2015; Gottfredson et al., 2015; Cheng et al., 2017; Al-Ubaydli et al., 2017).

The Al-Ubaydli et al. model changes the discussion from one that exclusively focuses on the benefit side (voltage effect literature) to a broader metric that includes benefits and costs (BC). If BC changes at scale—meaning benefits and/or costs change as scale changes—it is a manifestation of the scale-up effect.

Put another way, implementation science and the science of scaling are synonyms, or umbrellas to define this area of inquiry. The voltage effect was defined as a benefit-side

consideration by implementation scientists around 2007. We broadened the voltage effect to include the cost side in our science of scaling work, and denoted that as the scale up effect. In this way, we added an economic model as one approach under the broad umbrella of the science of scaling, or the science of using science.

This evolution to focus on costs as well as benefits is an important one because many policymakers are attracted to policies that they expect will provide the greatest benefit to the population within time, money, and resource constraints.

Additionally, the implementation science literature primarily focuses on fidelity as a key to addressing the voltage problem. While fidelity of the original research study at scale is certainly important, the richness of the economic environment surrounding most interventions calls for a more holistic approach to studying the scale-up problem. Thus, the economic model identifies four types of threats to scalability to be aware of.

What are the objectives and limitations of modeling?

Economic models provide theoretical frameworks for understanding complex systems. Like all economic models, this model of the science of scaling has limitations. It is built on assumptions that may or may not be true in every situation, and by necessity it includes some players and some elements but not others.

Consider this: The chain that connects scientific research discovery and evidence-based policy enactment contains three primary links: a) funding basic research, b) providing the knowledge creation market with the optimal incentives for researchers to design, implement, and report scientific results, and c) developing a system whereby policy makers have the appropriate incentives to adopt effective policies, and to implement those policies with rigorous evaluation methods to ensure continual improvement.

The economic model behind the science of scaling is an abstraction that focuses on the second link while leaving out the other two links. This does not mean that the other two links are irrelevant. On the contrary, one weak link will prevent the entire system from achieving its promise. This particular framework helps us understand how to ensure that the process by which the knowledge creation market provides incentives to researchers is not that weak link.

One particularly notable element that is omitted from this model (because it is part of the third link in the chain) is that of political will. In most cases, scaling of an intervention simply isn't possible without it. Additionally, while the model highlights many theoretical reasons why the scale-up effect may occur as programs are scaled, empirical work must be completed to determine which pieces have empirical relevance.

Despite its limitations, this model introduces a logical and coherent economic framework to a field that, to date, has largely been devoid of economics. Given the importance of engaging all stakeholders and all disciplines in addressing the scale-up effect, this is a critical contribution.

Who stands to benefit from advances in the science of scaling?

The scale-up problem can lead to a vast waste of resources, a missed opportunity to improve people's lives, and a diminution in the public's trust in the role of science in the policymaking process. For these reasons, everyone stands to benefit from a better understanding of scaling.

Researchers who apply the science of scaling when crafting their research designs will produce results that are more useful in the real world and ultimately more impactful. Policymakers who evaluate programs and policies through the lens of the scale-up effect will adopt those that are more likely to scale up effectively in their local context, producing better outcomes for their constituents, and, as a result, likely benefitting electorally. Members of the public will, of course reap benefits in the form of more effective (and cost-effective) public policies.

Why was the science of scaling chosen as the focus of the inaugural year of the Griffin Applied Economics Incubator?

The science of scaling represents the next frontier—a critically important one—in evidence-based policymaking. It is the missing link in the relationship between research and policy; the part that leads one to ask, "When this intervention is expanded to broader and larger populations, can we expect the same level of efficacy that we observed in the small-scale setting? And if not, what can we do differently?" Heretofore, researchers have focused almost entirely on generating evidence of intervention effects and testing the theories behind them. Yet, by failing to consider how such programs will be implemented at scale from the very beginning of the research design, critical questions go unasked and unanswered. And the utility of the research is greatly diminished.

Given the Griffin Applied Economics Incubator's emphasis on driving broad-based thought and policy changes, and establishing University of Chicago as a hub for doing so, focusing on an issue that stands to profoundly enhance the knowledge-creation and policymaking processes was an obvious choice.

Why is the Incubator exploring the science of scaling through the lens of early childhood specifically?

Early childhood was chosen as the lens through which the Incubator would explore the science of scaling for three reasons. First and foremost, because, developmentally, this is the most important period in an individual's life. Effective scaling of interventions that promote development in the early years has the potential to lift millions of children out of poverty. Second, the early childhood system in the U.S. is complex and notoriously fragmented, making it particularly challenging to study, implement, and scale programs in this field. It is our hope that examining a particularly vexing example will prove most informative to all, allowing us to draw lessons that apply to all public policy domains. Finally, early childhood was a natural choice because it is the domain of the TMW Center for Early Learning + Public Health, a center that has been doing groundbreaking work related to scaling, and the principal

What are the objectives of the Griffin Applied Economics Incubator's year focused on the science of scaling?

The idea behind the Incubator is to bring together as many of the world's brightest minds as possible to work on a critical issue, to encourage those individuals to pursue research on the topic, to collaborate with one another, to inspire young scholars, and to build momentum on the path forward.

During year one, if we can link research from different disciplines and connect thought leaders from economics, psychology, sociology, medicine and public policy to stimulate new ideas

that develop innovative approaches in early childhood and the science of scaling, then we will have met our goal.