

Why Every Organization Needs a Scale Unit

By John A. List and Dana Suskind

While President Biden spent much of his first State of the Union address on the economy, the pandemic, and Putin's invasion of the Ukraine, he also discussed another urgent issue: the plight of American families.

The president laid out his plan for easing the financial strain on middle- and working- class people who want to give their children the best possible future. One of the levers Biden hopes to pull is to make pre-k for three- and four-year-olds free. As the co-directors of the University of Chicago's Center for Early Learning + Public Health, we applaud him. But to make this plan and others like it a success, he'll need more than just political dexterity to push policies through the House and Senate. He'll need the latest science on how to scale government programs.

To many Americans, free pre-k might seem like a no-brainer policy for a country with so many overworked, underpaid parents. But critics point out that the benefits of similar initiatives in the past have not always been evenly distributed demographically. In a few prior cases, some children showed more developmental gains than others; one such pre-k program in Tennessee even appeared to hurt long-term outcomes for some students. This is discouraging, but to invoke past results like these misses the point. We shouldn't be having a zero-sum discussion about whether or not to make investments in support of families. Instead, we should be figuring out how to design programs that work across different regions and socioeconomic groups, within budgetary constraints. Of course, the need for the effective scaling of social policies goes beyond just universal pre-k. It applies to all policies aimed at making people's lives better. But how do we ensure they work? We have a plan.

As social scientists eager to see taxpayer money have maximum positive impact across society, we have studied why programs with great potential in pilot studies often fall apart after expansion. This phenomenon is known as a "voltage drop," or the scale-up problem. It afflicts numerous promising policy initiatives, not just free pre-k; everything from programs aimed at reducing energy consumption to promoting diversity and inclusiveness. When a program funded by state or federal money fails to accomplish its objectives when scaled, it is easy to understand the reluctance of many lawmakers to support others like it. Our solution to this problem is simple: every organization should have Scale Units.

In the wake of Richard Thaler and Cass Sunstein's bestseller *Nudge*, numerous companies and governments around the world have adopted Nudge Units. These are groups that use findings from research on human behavior to "nudge" people toward making certain decisions. Such behavioral insight teams help spur citizens to vote and pay their taxes, for example. We propose the formation of similar groups for designing and analyzing social programs, except the purpose of Scale Units is to determine whether a seemingly good idea is in fact scalable in practice. We have identified the five vital signs that all scalable enterprises must possess. In other words, programs worth scaling clear the following five hurdles.

The first cause of failure at scale is false positives, cases where the program didn't actually work in the first place, though it seemed otherwise, either because of a statistical error or because researchers massaged results to create the appearance of success.

The second is overestimating how much of the population your idea benefits. Often this is the result of assuming that the small subset of people affected in a pilot program are more representative of the general population than they actually are. We ourselves faced this problem when a curriculum we designed for an experimental pre-school in Chicago Heights had more success with Hispanic children than white children. This meant it wouldn't scale universally.

The third hurdle is failing to evaluate whether your initial success depends on unscalable ingredients, or unique circumstances that can't be replicated at scale. For example, when smart home thermostats were touted as the next big thing in fighting climate change, the original proponents didn't forecast that when rolled out humans were the end users. Such users were not like the automata who gave the initial optimistic estimates, but these were users who had built-in biases that led them to fiddle with the default settings. The result? A massive failure at scale: no energy was conserved.

The fourth is when the implementation of a program has unintended consequences, or spillovers, that backfire against that very program. Imagine an upskilling initiative that retrain unemployed workers—only to flood the local labor market so few of these people obtain a job with their new skills.

The fifth and last obstacle all scalable programs must pass is the “supply-side economics” of scaling. In other words, do the societal benefits outweigh the fiscal costs? When investing in our children and our nation's future we must be sure to implement programs that don't cause voltage drops because of diseconomies of scale: as we expand, it becomes more and more expensive to hire the necessary inputs.

We envision Scale Units as seamlessly working within an organization to unswervingly inspect and guide decision making around the scalability of programs, and whether the idea possesses the five vital signs. Once the presence of the five vital signs is confirmed, the likelihood of success skyrockets. With Scale Units, the passing of social policies that benefit Americans can be less political and more scientific—and we can use our nation's wealth in ways that yield big results.