

The science of scale: what works to implement effective education programs?



by Noam Angrist, Penny Morton, Andrew Leigh, Filemon Ray Javier and Eleanor Williams

18 March 2026

Evidence from education systems around the world shows that success at pilot stage does not guarantee success at scale. On 4 December 2025, a [keynote panel](#) at the Australasian Aid and International Development Conference was brought together to discuss the science of scale: how we can work together to ensure programs remain effective as they grow. The discussion drew on global evidence, including work linked to the [What Works Hub for Global Education](#) (WWHGE).

The panel opened with a shared recognition that education and development policy requires humility. As noted by Andrew Leigh, Australia's Assistant Minister for Productivity, Competition, Charities and Treasury in the Australian Government, theory alone is rarely sufficient in complex systems, and rigorous evidence is essential to understanding what works, particularly when programs are scaled up.

Policies built on untested assumptions can fail without evidence. Leigh illustrated this with an example from public health. In the early 2000s, he recalled, William Easterly argued that giving out insecticide-treated bed nets to prevent malaria would not work because people would not use them.

“A series of randomised trials tested this claim and the results were utterly conclusive,” he said. When people were asked to pay for bed nets, uptake fell; when nets were offered free, use increased. “The randomised trials changed policy,” he explained, with free bed nets being rolled out across Africa. It was “a high-quality evaluation [that] had the effect of saving hundreds of thousands of children’s lives”.

The discussion focused not only on identifying effective interventions, but on how collaboration between policymakers, implementers and evaluators can help sustain impact as programs expand.

Noam Angrist, Academic Director of WWHGE, highlighted the critical role of

evidence in identifying what works to improve education outcomes in low- and middle-income countries. WWHGE has found that pedagogical approaches — such as using [diagnostic assessments to target instruction to a child's level](#) — are among the most reliable ways to improve learning outcomes. Over half a dozen randomised evaluations across countries and contexts demonstrate strong impacts. Dedicating resources to these types of evidence-backed policies can help ensure good value for money.

Eleanor Williams, Managing Director at the Australian Centre for Evaluation, emphasised that these insights matter most when evaluation is closely linked to delivery. “We need evaluators to work alongside policy and implementation professionals,” she said, warning that scaling is too often treated “like it’s a dark art”, despite [a growing body of literature and experience](#) on how to scale evidence-based programs effectively.

It was also clear how visible political leadership can accelerate implementation and motivate frontline educators. Beyond technical design and evaluation, the panel highlighted the role of political leadership in translating policy into practice. Visible commitment from senior leaders can send a powerful signal throughout education systems, reinforcing the importance of reforms and sustaining momentum during implementation.

Filemon Ray Javier, Undersecretary for Legal and Legislative Affairs in the Philippines Department of Education, described how the Secretary of Education recently assigned senior policymakers to regions to support the implementation of the [Academic Recovery and Accessible Learning](#) law focused on improving learning outcomes. The public saw undersecretaries and assistant secretaries crossing rivers and trekking through mountains to reach remote schools. “It shows that even the people in the government care about this program,” he said. In just three months, he reported, there had been a 16% increase in the number of learners meeting grade-level standards. “Imagine what we can do in the next three years,” he added.

When leadership engagement is paired with clear accountability and support for implementation, it can help overcome common challenges in scaling, from uneven uptake to implementation fatigue. Political buy-in is a critical enabler of scale.

The panel discussed how evaluation must evolve as programs scale, shaping what should be adapted and testing what must be retained. Panellists underscored that evaluation is not static: the questions that matter at pilot stage are not the same as those that matter once a program begins to scale. Early evaluations often focus on whether an intervention works, while later stages require clarity on which elements

are most essential for cost-effectiveness and which can be adapted across local contexts while preserving impact.

Penny Morton, former Minister-Counsellor at the Australian High Commission in PNG, reflected on this progression from an implementation perspective. She outlined how the [Department of Foreign Affairs and Trade \(DFAT\)](#) has focused on establishing a genuine partnership with the PNG Government, providing bespoke technical support for country-led reforms such as the scaling-up of structured pedagogy. “Improving the education outcomes at the foundational level is hard work, requiring long term investments and patience,” she said.

Eleanor Williams also commented on the evolution of evaluation in pilots to scale. “There is a journey you go on as an evaluation unit to understand how to deliver context-appropriate evaluation,” she said. Moving from two sites to four sites is the point at which programs need to identify which components are fixed and which require local adaptation. Scaling further — “from four sites to a thousand sites” — represents a different level of maturity, requiring evaluators with the skills to recognise what type of evaluation is needed at each stage.

[A/B testing](#) is one practical tool pioneered by the technology sector which is now taking off in various social sectors. By comparing two versions of a program rather than testing everything against a pure control group, A/B testing keeps the power of randomisation while being embedded in implementation at scale, with everyone getting a version of the policy or program. A/B testing is characterised by “Three Rs” — Rigorous, Rapid and Regular — ensuring results are generated in real-time, informing and iteratively improving programs as they scale.

A recurring theme of the discussion was the opportunity to reduce the cost and complexity of evaluation by making better use of existing administrative and system-level data. As data systems improve, administrative data offer powerful opportunities to track outcomes at scale without relying solely on bespoke surveys or costly new data collection. As Noam Angrist highlighted, “Evaluation doesn’t have to involve expensive data collection. Much can be learned from data that already exists within education systems.”

Finally, the panel discussed how careful tracking of implementation fidelity, including through real-time monitoring, helps systems evolve. The panel stressed that scaling effective education programs depends on continuous feedback once programs are underway. Regular monitoring creates the conditions for adaptation, allowing systems to identify who is benefiting — and who is being left behind — and to respond accordingly.

DEVPOLICYBLOG

Describing the Philippines' experience, Filemon Ray Javier explained how this plays out in practice. "We do regular, periodic assessments among our learners," he said, noting that previously there had been little systematic testing. "Assessments are now conducted at the beginning, middle and end of the school year, allowing results to be compared and programs adjusted. Crucially, this monitoring made it possible to ask why some learners were not improving."

The science of implementation and scale is finally taking off, and with it, improved outcomes for millions who deserve evidence-backed, well-implemented policies and programs.

You can watch a recording of the panel and other 2025 Australasian AID Conference sessions on [Devpolicy's YouTube channel](#).

Author/s:

Noam Angrist

Noam Angrist is the Academic Director for the What Works for Global Education Hub at the University of Oxford.

Penny Morton

Penny Morton is currently Assistant Secretary in the Multilateral Policy Branch at Australia's Department of Foreign Affairs and Trade. She was previously Minister-Counsellor at the Australian High Commission in Papua New Guinea.

Andrew Leigh

Dr Andrew Leigh is the Assistant Minister for Productivity, Competition, Charities and Treasury, and Federal Member for Fenner in the ACT. Prior to being elected in 2010, Leigh was a professor of economics at the Australian National University.

Filemon Ray Javier

Filemon Ray Javier is the Legal and Legislative Affairs Undersecretary in the Department of Education in The Philippines.

Eleanor Williams

Eleanor Williams is the Managing Director of the Australian Centre for Evaluation at the Department of Treasury.

Link:

<https://devpolicy.org/the-science-of-scale-what-works-to-implement-effective-education-programs-20260318/>