2025/26

A promising study is not a reliable predictor of success

We see a study with big results and think we've found the next breakthrough. But what we know about initial studies tells us to be more cautious.

This document is authored by Lee Bentley, Stephen Morris.

Produced and edited by Doris Hanappi PhD, Medea Fitzlaff MA, Miriam Daepp,
Jacobs Center for Productive Youth Development UZH

Funding Acknowledgement

This Policy Brief was supported by the European Commission action HORIZON-CL2-2023-TRANSFORMATIONS-01 for the project LEARN (Longitudinal Educational Achievements: Reducing iNequalities), Grant Agreement 101132531.

Funded by the European Union's Horizon Europe Programme HORIZON-CL2-2023-TRANSFORMATIONS-01 Grant Agreement 101132531 and co-funded by UK Research and Innovation (UKRI) grant agreement numbers 10107302, 10092247 and the Swiss State Secretariat for Education, Research and Innovation (SERI)

LONGITUDINAL EDUCATIONAL ACHIEVEMENTS: REDUCING INEQUALITIES
LEARN

When a new educational intervention shows a large, positive effect, there can be a strong impulse to scale it up quickly. This belief assumes that promising early results are a reliable predictor of success. However, this assumption mistakes a promising start for a predictable finish, risking significant public investment in interventions that are not yet proven at scale.

This is reinforced by the pressure to identify and fund “what works” solutions. The problem is related to the issues of publication bias, where studies showing positive effects are more likely to be published while those with no, or negative, findings remain in the “file drawer” (Franco, Malhotra, and Simonovits 2014). This can create a skewed evidence-base where the most optimistic findings are most visible, leading to scaling decisions based on an overly positive picture of effectiveness.

Many early trials are statistically “underpowered” due to small sample sizes (Lortie-Forgues and Inglis 2019). Such studies are not only more likely to miss a real effect, they are also prone to exaggerating the size of an effect when one is detected (Sims et al. 2022). This explains why impressive results often aren't replicated in further assessments, with large initial effect sizes diminishing, particularly when the first experiment was small.

Researchers have cautioned that underpowered studies “are prone to producing exaggerated effect sizes and false positives” (Sims et al. 2022). This was highlighted in our research, with one expert interviewed for the LEARN project observing the worrying trend of, “Huge evaluations, small effects when scaled up... Effects tend to reduce at each stage.” Awareness of these issues has grown, largely due to the 'replication crisis' in many academic disciplines. Requirements, such as pre-registering studies, and publication regardless of outcome are increasingly common. Yet the fundamental risks of early-stage research remain.

Practical challenges also make scaling difficult. An initial study often has ideal conditions: extra resources, motivated participants, and close oversight. A wider rollout must contend with far greater variability in school contexts, student populations, and implementation (Hofmann 2024). These real-world factors almost inevitably lead to a dilution of the intervention's original impact.

Therefore, the challenge for policy is not simply to identify promising studies, but to build a more patient and rigorous evaluation pipeline that also considers factors like cost and scalability (Kraft 2020). A staged approach—involving replication and testing in diverse, realistic settings—is required, rather than a leap from a single study to a national rollout. This model ensures programmes are robust and adaptable before major investment, shifting the goal from finding quick wins to building sustainable solutions ready for scale.

NOTE

Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

References

Franco, Annie, Neil Malhotra, and Gabor Simonovits. 2014. ‘Publication Bias in the Social Sciences: Unlocking the File Drawer’. *Science* 345(6203):1502–5. doi:10.1126/science.1255484.

Hofmann, Riikka. 2024. ‘The Four Paradoxes That Stop Practitioner from Using Research to Change Professional Practice and How to Overcome Them’. *Education Sciences* 14(9):996. doi:10.3390/educsci14090996.

Kraft, Matthew A. 2020. ‘Interpreting Effect Sizes of Education Interventions’. *Educational Researcher* 49(4):241–53. doi:10.3102/0013189X20912798.

Lortie-Forgues, Hugues, and Matthew Inglis. 2019. ‘Rigorous Large-Scale Educational RCTs Are Often Uninformative: Should We Be Concerned?’ *Educational Researcher* 48(3):158–66.

Sims, Sam, Jake Anders, Matthew Inglis, and Hugues Lortie-Forgues. 2022. ‘Quantifying “Promising Trials Bias” in Randomized Controlled Trials in Education’. *Journal of Research on Educational Effectiveness.* doi:10.1080/19345747.2022.2090470.