

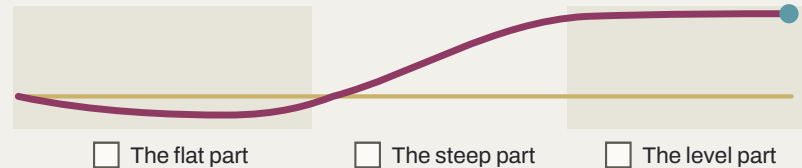
Where is your program on the curve?

For a team that includes the leader who will have to defend the decision. Plan on one meeting, and bring the data you already have.

1 Name the program and the outcome it is meant to move.

Then say how the program reaches that outcome. If the path takes more than two steps to describe, write that down too.

2 Mark where the program sits now, and what puts it there.



WHAT PUTS IT THERE

☐ We can't tell

If the team can't tell, the baseline is the place to start.

3 Answer the three questions for that position.

What matters most to know right now?

Flat: is it happening as designed? Steep: who are the gains reaching? Level: have the gains stopped, or only slowed?

What do we need to keep learning over time?

Flat: a baseline and the measures you will repeat. Steep: the same measures, collected the same way. Level: which parts are still doing the work, and which your organization could stop paying for.

Which data are requirement-driven rather than value-driven?

Name the numbers the next renewal decision will use, and what you will show next to them.

4 Run a pre-mortem.

Imagine it is a year from now and the decision about this program turned out wrong, whether your organization cut it too early or kept paying after the gains stopped. Write the most likely reason. Then imagine the opposite: a year from now, the program is working. Write what the team did to get it there.

IT WENT WRONG BECAUSE

IT WORKED BECAUSE THE TEAM

5 Before you leave the room.

Name the one measure that would have warned you, who owns it, and when the team will look at it again.

THE MEASURE

WHO OWNS IT

WHEN WE LOOK AGAIN

If you want support working out where your program sits, I would love to hear about it.



What the research says

The research behind this worksheet. The curve is a teaching picture. Research supports the slow start and the gains that build with time, and real programs vary in how long each part lasts. Links to every source are at swerdlowpartners.com/the-curve.

Programs often take two to four years to run as designed.

Fixsen and colleagues reviewed implementation research across education, mental health, and social services and found that getting a program fully running in a new site often takes two to four years. Most evaluations happen before that point, so a program can look ineffective because it was measured too early. Fullan calls the early slump the implementation dip.

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature*. National Implementation Research Network, University of South Florida.

Fullan, M. (2001). *Leading in a culture of change*. Jossey-Bass.

Reform effects tend to grow with years of implementation.

Borman and colleagues' meta-analysis of 29 comprehensive school reform models found particularly strong effects in schools that had implemented a model for five years or more. May and Supovitz followed one reform model in Rochester, New York, for 11 years and found yearly effects in elementary and middle grades reading and math that kept adding up.

Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research*, 73(2). doi.org/10.3102/00346543073002125

May, H., & Supovitz, J. A. (2006). Capturing the cumulative effects of school reform: An 11-year study of the impacts of America's Choice on student achievement. *Educational Evaluation and Policy Analysis*, 28(3). doi.org/10.3102/01623737028003231

Programs often continue only in part.

Wiltsey Stirman and colleagues reviewed studies of what happens to new programs after initial support ends. Partial continuation was more common than continuing the whole program, and where fidelity was rated independently, fewer than half of the providers kept the practice up at a high level.

Wiltsey Stirman, S., Kimberly, J., Cook, N., Calloway, A., Castro, F., & Charns, M. (2012). The sustainability of new programs and innovations. *Implementation Science*, 7, 17. doi.org/10.1186/1748-5908-7-17

Short leadership tenures feed a cycle of adopting and dropping reforms.

In his study of urban districts, Hess describes policy churn. Superintendents, whose tenure is usually three years or less, face pressure to show they are making a difference, so new reforms keep arriving, and the churn seldom produces lasting change in classrooms.

Hess, F. M. (1998). *Spinning wheels: The politics of urban school reform*. Brookings Institution Press.

How well a program is implemented shapes what it achieves.

Durlak and DuPre's review of implementation research found mean effect sizes two to three times higher in programs that were implemented well. Positive results often came with implementation levels around 60 percent, and few studies reached levels above 80 percent.

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters. *American Journal of Community Psychology*, 41(3–4), 327–350. doi.org/10.1007/s10464-008-9165-0

Measurement for improvement starts with a baseline.

Solberg, Mosser, and McDonald separate measurement for improvement from measurement for accountability and for research. Measurement for improvement includes establishing a baseline and checking newly changed processes, which is why the flat part starts with a baseline recorded before anything changes.

Solberg, L. I., Mosser, G., & McDonald, S. (1997). The three faces of performance measurement: Improvement, accountability, and research. *Joint Commission Journal on Quality Improvement*, 23(3), 135–147.



What the research says, continued

Required data often can't answer improvement questions.

Solberg and colleagues write that data for accountability “do not usually illuminate how the outcomes were achieved or how processes might be changed to improve them.” In the RAND studies, state test results generally arrived too late to adjust the current school year, and more than 80 percent of superintendents found local assessments more useful for decisions than state tests.

Solberg, Mosser, & McDonald (1997), above.

Marsh, J. A., Pane, J. F., & Hamilton, L. S. (2006). *Making sense of data-driven decision making in education: Evidence from recent RAND research*. RAND Corporation.

Imagining a failure in advance helps teams find its causes.

Klein developed the pre-mortem and reports that research by Mitchell, Russo, and Pennington found that imagining an event had already happened increased people's ability to identify reasons for future outcomes by 30 percent. The worksheet asks the question both ways, which adapts Klein's method.

Klein, G. (2007, September). Performing a project premortem. *Harvard Business Review*.

If you want support working out where your program sits, I would love to hear about it.

