On Scaling Instructional Improvement

A synthesis of The Coaching Letter #s 231, 232, 233, 234 & 235

This is an argument for thinking differently about scale from the "typical" approach in most school systems, because the typical approach frequently misdiagnoses the nature of teaching. It assumes that effective instructional practice can be transferred primarily through rollout, exposure, compliance, and replication, when teaching is actually adaptive, context-dependent, and deeply shaped by tacit knowledge. Not all variation is bad. Systems need adaptive expertise and local responsiveness. But when variation means that students are not guaranteed access to ambitious instruction, variation becomes an equity problem. Therefore, school systems need to be adept at scaling effective instructional practice.

1. Most school systems operate from a technical theory of scale

Most educational leaders think about scale as spread, measured in counts:

- more teachers trained
- more classrooms implementing
- more materials distributed

We call this **Logic Alpha**. It treats scale as a technical rollout problem and values fidelity, implementation, standardization, accountability, and consistency. The vocabulary is *turnkey*, *launch*, *rollout*, and *buy-in*.
Logic Alpha is the right approach when the work is genuinely **technical**—fire drills, food service, transportation, budgeting. But most instructional improvement is not technical. It is **adaptive**. Drawing on Heifetz and Linsky: technical problems can be solved *for* people, because someone already knows what to do. Adaptive challenges require changes in beliefs, habits, and ways of working, and can only be addressed *with* people. Applying a technical logic to an adaptive challenge produces compliance, not learning.

Logic Phi asks a different question: how do we create the conditions under which effective practices can take root, deepen, adapt, and propagate over time? It treats scale not as quantitative spread but as multidimensional. Drawing on Cynthia Coburn (2003), scale has four dimensions:

- **Depth** — changes in teachers' underlying assumptions about students, content, and what counts as good instruction.
- **Sustainability** — durability of the change beyond the initial push.
- **Spread** — diffusion of the norms, beliefs, and principles behind the practice, not just the visible practice itself.
- **Shift in reform ownership** — the system can sustain the work without constant external oversight or outside expertise.

Logic Phi is not structureless. Drawing on Karl Weick's distinction between tight and loose coupling, it is **tight on a small number of core practices and loose on how, where, and through whom the work spreads**. That is not a compromise between top-down and bottom-up. It is a design principle.

> ***Therefore,*** *before deciding how to scale, leaders need to diagnose whether the work in front of them is technical or adaptive—and design accordingly.*

---

2. The biggest challenge is confusing form and function.

Most instructional improvement efforts scale form rather than function. Systems reproduce the visible form while losing the underlying mechanism. Classrooms can look similar on the surface while students experience very different intellectual work—and leaders can plausibly claim both that the practice was implemented and that it didn't work.

Form is what a practice *looks like*:

- students in groups
- students on task
- teacher circulating
- student discourse

Function is what the practice is *doing*:

- maintaining cognitive demand and supporting sensemaking
- making thinking visible and generating feedback
- cultivating academic and social belonging and perceived self-efficacy

A theory of action must specify not merely what people do, but *why* those actions are expected to produce improved learning. It needs a ***because***—a research-grounded mechanism connecting action to outcome. Without one, the system is scaling a theory of form, not a theory of action.

> ***Therefore,*** *the question driving any scaling effort should be: what is the mechanism, and what evidence do we have that it works?*

---

3. Teaching is hard to scale because it relies on tacit knowledge.

Teaching contains enormous amounts of **tacit knowledge**—what experts know that they cannot fully articulate. This includes:

- judgment
- perception
- pattern recognition
- situational awareness
- adaptation under uncertainty

Expert teachers know more than they can say. But this does not mean teaching is an art in the sense of unique self-expression. It is closer to surgery, aviation, firefighting, or trial law, in which:

- shared standards matter
- expertise develops through deliberate practice
- adaptation is constant
- judgment is critical
- outcomes are life-changing

Two framings have to be rejected at once. The first says teaching is purely technical and can be scaled through scripts and checklists. The second says teaching is an art and therefore cannot be codified, improved collectively, or scaled. Both lead nowhere. The existence of uncertainty does not mean expertise is impossible—it means expertise operates probabilistically rather than mechanically.

> ***Therefore,*** *the tacit nature of instructional expertise does not exempt the profession from the systematic study, codification, and testing of effective practice. It makes those moves more important.*

---

4. Instructional improvement depends on making expert practice visible and learnable.

If teaching expertise is tacit in significant ways, then the work of instructional improvement is making powerful forms of practice **increasingly visible, discussable, testable, and learnable**—without flattening them into scripts. Practices change because people try them and experience success. The lever is creating the conditions under which people will try.

A practical approach:

1. **Start from positive outliers.** Identify where the practice is already alive in the system. Depth, not seniority, is the criterion. These are the bright spots that show the work is possible here, in this context.
2. **Codify the practice as a recipe.** Recipes are testable articulations of high-leverage practice—not scripts, checklists, or teacher-proofing. They name the moves clearly enough that a team can try them and learn from the result.
3. **Make the team the unit of learning.** Recipes allow teacher teams to test practices, adapt them, study impact, and refine collective understanding. The team becomes a researcher of its own practice.
4. **"Double the Sarahs."** Find the people in the system who get the work, support them directly, and enlist them in finding others who will do the same. Leadership is not synonymous with title.

The goal is not uniformity for its own sake. The goal is reducing harmful variation in cognitive demand, expectations, and access to grade-level work.

> ***Therefore,*** *there is enormous leverage in building the structures and relationships that let teams become researchers of their own practice.*

---

5. Pre-mortems can help forestall failure in scaling.

Even the best thought-out attempts at improving a system are fraught with potential problems, because systems are complex. This is especially true when the work being scaled is adaptive, tacit, and context-dependent. That is why pre-mortems are so valuable. A pre-mortem, developed by Gary Klein, asks teams to imagine that an initiative has failed in the future and then work backward to identify the reasons why. Unlike ordinary planning processes, which often assume success and focus on rollout logistics, pre-mortems assume that systems will respond unpredictably and that interventions will create unintended consequences.

In instructional improvement, this matters enormously because many scaling failures are not failures of intention. They are failures of interpretation, architecture, incentives, or adaptation. For example:

- A cognitively demanding instructional routine becomes reduced to superficial compliance with visible structures.
- Coaching becomes consumed by supporting novice teachers created by the district's own staffing decisions.
- Collaborative planning time becomes logistical coordination rather than collective inquiry into teaching and learning.
- A successful pilot fails at scale because the pilot depended on unusual expertise, unusually favorable conditions, or unsustainably high levels of support.
- Teachers reproduce the visible form of a practice while losing the underlying instructional function.

Pre-mortems are exercises in systems thinking. They force leaders to ask:

- What assumptions are we making?
- What conditions does this practice depend on?
- What variation might emerge as this spreads?
- What would cosmetic implementation look like?
- What incentives might distort the work?
- What mental models could reinterpret the practice in ways that undermine its purpose?

The more tacit and adaptive the work, the more important this kind of prospective thinking becomes. Technical work can often survive imperfect implementation. Adaptive work is much more vulnerable to distortion because its success depends on interpretation, judgment, relationships, and shared understanding.

> ***Therefore, ,*** *leaders should treat scaling not as a rollout problem to be managed, but as an ongoing process of anticipating how the system will respond and accounting for that.*

6. Mental models are the highest-leverage point.

Systems respond to interventions through the mental models of the people inside them. Educators interpret rigor, support, engagement, scaffolding, and "good teaching" through their existing assumptions and beliefs. As a result, identical initiatives produce different outcomes in different systems. **Implementation is always interpretation.**

Donella Meadows argued that the highest-leverage place to intervene in a system is the mental models that allow the system to reproduce itself. In instructional improvement, those models include beliefs about:

- what students are capable of
- what counts as equitable and ambitious instruction
- the role of teacher teams in testing and refining practice
- the practice architectures that support ambitious teaching
- the role of leadership in creating the conditions under which better teaching can emerge

Mental models rarely shift through persuasion alone. They shift through:

- enacted experience
- collective inquiry
- evidence from practice
- seeing students succeed in new ways

Belief and practice form a reinforcing loop. Changing one without the other doesn't hold.

> ***Therefore,*** *the work of scaling instructional improvement is not spreading programs. It is creating the conditions under which increasingly powerful forms of practice, understanding, and collective learning propagate through a system—and the people inside it come to see the work differently.*