

# The Scaling Field Guide *Companion*

*An evidence base for deciding whether, how,  
and when to scale an innovation.*

**01**

## **Scalability**

*Should we scale?*

---

**02**

## **Scaling**

*How will we scale?*

---

**03**

## **Scale**

*Is the innovation working  
at scale?*

**Matt Healey**

Principal Consultant

APRIL 2026

[fpconsulting.com.au/scaling-field-guide](https://fpconsulting.com.au/scaling-field-guide)



---

## 01 Introduction

*Three lenses in a complex system*

---

## 02 Scalability: *Should we scale?*

---

2.1	Definitions: scalability, scaling, and spread .....	7
2.2	Should we scale? Justification as a principle .....	8
2.3	The Scaling Focus: what is actually being scaled? .....	9
2.4	Directions for scaling .....	10
2.5	Optimal scale and trade-offs .....	12
2.6	From evidence to tool: the O-DEAR questions .....	12

## 03 Scaling: *How will we scale?*

---

3.1	Scaling in a complex system .....	16
3.2	Scaling Pathways .....	17
3.3	Strategic considerations: what makes scaling work .....	18
3.4	From evidence to tool: the Scaling Canvas .....	20

## 04 Scale: *Is it working at scale?*

---

4.1	The fidelity-adaptation tension .....	23
4.2	Implementation drivers at scale .....	24
4.3	Dynamic evaluation .....	24
4.4	Sustainability as continued benefit at scale .....	25
4.5	From evidence to tool: the Fidelity Checker .....	26

## 05 References & Appendix A: Glossary

---

*Scaling is a deliberate practice in a system that does not necessarily want to be changed.*

# 01

---

SECTION ONE

## **Three lenses in a complex system**

---

**S**caling means different things in different fields. In public health, it often means reaching a greater share of the eligible population while holding effectiveness constant. In social innovation, it can mean changing policy, shifting cultural norms, or building a delivery network that carries a core idea without central control. In international development, it has meant everything from replicating pilot programs to institutionalising new ways of working across government.

The scaling literature isn't one literature. It's several, with different assumptions and vocabularies, overlapping at the edges. That variation isn't a problem to be solved by picking one definition. It's the reality anyone scaling an innovation is working in. Whatever your innovation, the path to scale runs through other people's systems (regulators, funders, employers, service users, communities), each with their own logic. Scaling is a deliberate practice in a system that doesn't necessarily want to be changed (Snowden & Boone, 2007; Westley, Zimmerman, & Patton, 2006; Hawe, Shiell, & Riley, 2009).

#### SOURCES

Milat et al., 2016; Moore et al., 2015; Fraser, 2023; Nesta, 2014; Cooley & Linn, 2014; Hartmann & Linn, 2008.

That framing does two things. First, it puts complexity at the start, where it belongs. Second, the structure follows the questions organisations actually face, not a stage model. The three lenses (Scalability, Scaling, and Scale) aren't sequential phases. They're ongoing questions, asked and re-asked as an innovation matures.

#### THE THREE LENSES

**Scalability.** *Should we scale?* A judgement about whether an innovation could be expanded while retaining effectiveness.

**Scaling.** *How will we scale?* The deliberate processes of introducing an innovation at wider reach through a delivery structure.

**Scale.** *Is the innovation working at scale?* The ongoing work of checking fidelity, adaptation, and continued benefit once at scale.



FIGURE 1. THE THREE LENSES, WITH THE FPC TOOL FOR EACH

The lenses don't sit side by side in practice. They unfold in time, and the move between them is rarely clean. Deciding to scale is a threshold you cross. Not a switch you flip. Scaling itself is the non-linear work of negotiating real-world conditions (funders, partners, delivery contexts) that don't arrange themselves in a straight line. Scale is the place you arrive at, and the thing you then have to keep evaluating. The shift is from "should we?" to "how will we?" to "is it still working?"

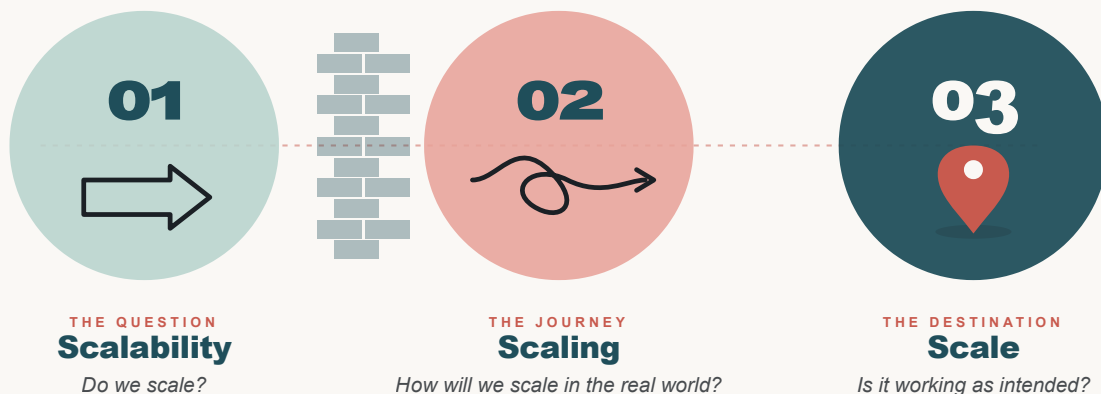


FIGURE 2. THEIR RELATIONSHIP: A QUESTION, A JOURNEY, A DESTINATION

PART ONE

*The first question isn't "how do we scale?" It's "should we?"*

TOOLS

*O-DEAR · Directions  
for Scaling · Scaling Focus*

02

---

SHOULD WE SCALE?

**Scalability**

*The first question for an organisation finishing a pilot isn't "how do we scale?" It's "should we?"*

That question is easily skipped, because the funding environment tends to reward ambitious answers (Bradach, 2003; Seelos & Mair, 2017). But scaling decisions made without careful thought about whether, what, and in which direction tend to founder. Either the innovation is scaled in a form that no longer works, or it is scaled past the point where it should have paused and tested again (Fixsen, Blase, & Fixsen, 2017; McLean & Gargani, 2019).

This part does four things. It sets out the definitions. It treats the ethics of "should we scale?" seriously. It introduces the concept of a Scaling Focus: what is actually being scaled. And it maps the directions an organisation can scale in. Each section closes with how the evidence feeds into the O-DEAR questions.

## 2.1 Definitions: scalability, scaling, and spread

The terminology in the scaling literature is loose. Some authors use *scaling* and *scaling up* interchangeably; others distinguish them. Some use *spread*, *diffusion*, and *dissemination* as if they mean the same thing; others treat each as a distinct process. A practical distinction is worth holding onto.

### WORKING DEFINITIONS

**Scalability.** The *potential* of an innovation to be scaled. A judgement about whether an intervention that has worked at small scale could be expanded under real-world conditions while retaining enough effectiveness to be worth the effort (Milat et al., 2013, 2016). Scalability is assessed; it is not achieved.

**Scaling.** The *deliberate process* of introducing an innovation through a delivery structure with the aim of reaching more of the eligible population and producing the intended benefits (Edwards, Riley, & Willis, 2017; Charif et al., 2017). Scaling is an active verb. It involves choice, effort, and judgement.

**Spread.** The movement of an idea or practice through a system, often without a central coordinator. Greenhalgh and colleagues' systematic review of innovation diffusion in service organisations (Greenhalgh et al., 2004) remains the canonical mapping. Bevan and colleagues (2024) make the useful point that spreading and scaling are often confused in practice. The two require different strategies and produce different outcomes, so the distinction is worth keeping.

An innovation can be spread without being scaled. Word of mouth and voluntary adoption move practices through a system without any of the deliberate work scaling requires. An innovation can also be scaled without being spread. A government program mandated by policy is scaled but rarely diffused in the organic sense. Treating these as separate processes helps clarify what an organisation is actually trying to do.

#### KEY DISTINCTION

An innovation can be spread without being scaled, and scaled without being spread. Treating them as separate processes helps clarify what an organisation is actually trying to do.

Milat and colleagues' (2013) scalability definition and Edwards and colleagues' (2017) definition of scaling are the two most widely used in the field. Alongside them sit a systematic stocktake of scalability assessment tools (Ben Charif et al., 2022), the ISAT instrument from the NSW Ministry of Health (Milat et al., 2020), and the Scale-Up Reflection Guide (Lee et al., 2023). Collectively, these confirm that the field has converged on a set of assessment dimensions broadly consistent with O-DEAR's scope, though no instrument has displaced qualitative, context-sensitive judgement as the core decision process.

The language in the broader literature has other variants worth knowing. *Scaling up* in the WHO and ExpandNet tradition describes deliberate efforts to increase the impact of successfully tested interventions to benefit more people and to embed them in policy on a lasting basis (Milat et al., 2016). *Functional scale-up* describes an innovation that piggy-backs on one that has already achieved widespread acceptance within a target population (WHO, 2010). *Scaling* in the Nesta lineage covers a broader set of processes (standardisation, diffusion, replication, proliferation) through which social innovations grow (Nesta, 2014). Appendix A consolidates these variants as a reference glossary.

## 2.2 Should we scale? Justification as a principle

Whether to scale is, first, a question about evidence and, second, a question about values. McLean and Gargani (2019) call the combination **Justification**, and they separate the two strands.

*Scaling is a choice that must be justified: by evidence that it can work, and by values that say it should.*

— MCLEAN & GARGANI, 2019

*Technical justification* rests on evidence that the innovation can produce the intended benefits in the proposed settings. If the pilot produced effects in a narrow set of conditions, there should be at least reasonable grounds to believe those effects will hold up when the

conditions widen. Bradach (2003) pushed program directors to ask whether replication was "reasonable and responsible". The test kept technical evidence honest by not letting ambition outrun it.

*Moral justification* goes further. It asks who decides, who benefits, and who bears the risk. An innovation may be technically justified and still not morally justified if the people who will be on the receiving end do not share the vision, or if the power dynamics around the scaling effort mean that the risks fall disproportionately on those with the least voice. Moral justification draws on evidence, but it is ultimately a judgement about values, and it is shared between the innovator and those most affected (McLean & Gargani, 2019).

Both strands matter, and they need each other. Technical evidence on its own can produce scaling efforts that are efficient but contested, or that reach the people who need them least. Good intentions on their own can produce well-meaning efforts that don't work. In practice, most scaling decisions sit somewhere with both kinds of evidence ambiguous; the task is to name the uncertainty honestly, not manufacture confidence.

#### PRINCIPLE

**Justification.** The first of McLean & Gargani's four guiding principles for scaling. Holds that scaling is a choice that must be justified on both technical and moral grounds.

This framing is consistent with Kohlhoff and colleagues' 2026 finding that successful scaling of early parenting interventions rests on a clear business model, genuine stakeholder buy-in, and high-quality evidence. They found this combination was rarely reducible to any one of these strands alone (Kohlhoff, Donkers, & Cibralic, 2026). It is also consistent with Fraser's (2023) reminder that "sometimes bigger isn't better": Justification can rule out scaling, not just rule it in.

## 2.3 The Scaling Focus: what is actually being scaled?

One of the clearer findings across the scaling literature is that organisations often do not have a shared answer to the question "what are we actually scaling?" Nesta (2014) called this the *core* of the innovation. Barker and colleagues (2016) used the term *scalable unit*. Edwards and colleagues (2017) referred to the *unit of scale-up*. Dees and colleagues (2004) framed it as a choice between scaling a *program*, an *organisation*, or a *principle*. Each is naming the same problem: if you cannot specify what is being scaled, you cannot make a sensible decision about how.

I prefer the term **Scaling Focus**. It's plain enough to function as a workshop prompt: "what are you focusing the scaling on?" And it holds the three sub-frames Nesta set out well.

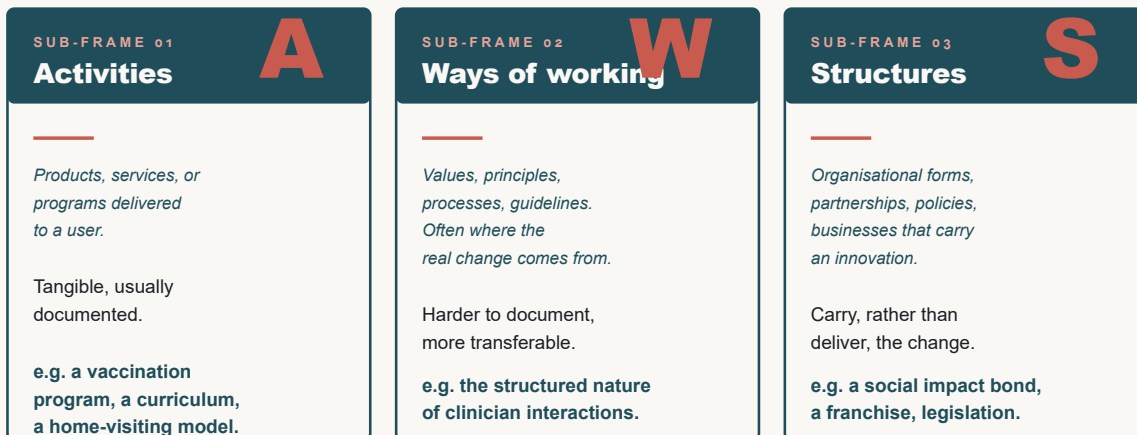


FIGURE 3. THREE SUB-FRAMES FOR SCALING FOCUS (AFTER NESTA, 2014)

An innovation often has more than one candidate Scaling Focus, and the choice of focus has large consequences. A program-as-Focus will typically require resourcing, training, fidelity monitoring, and a delivery structure. A principle-as-Focus can be scaled via advocacy, publishing, and training others, at lower cost but with less control. A structure-as-Focus can be scaled by influencing policy or forming alliances, often over longer timeframes. The Focus shapes the Direction (Section 2.4) and the Pathway (Part Two). They tend to come together; the choice of one constrains the others.

One further point. The Focus isn't fixed. Many of the innovations Nesta documented shifted their Focus as they scaled. A program turned out to have a transferable set of principles, or a set of principles turned out to need a delivery structure to land reliably. The Focus should be reviewed as part of ongoing dynamic evaluation (Part Three), not set once and defended (Fixsen et al., 2017).

## 2.4 Directions for scaling

Several typologies exist for describing the directions an innovation can scale in. I use five here. Each captures something the others miss.

### Scale out, scale up, scale deep

Originally developed for the non-profit sector by Moore, Riddell, and Vocisano (2015) and Westley and colleagues (2014), this typology has become the common shorthand in the social innovation field. *Scaling out* means reaching more people or places: replication, dissemination, spreading principles with adaptation. It is the default when "scaling" is used without qualifier, and often the weakest route by itself. Moore and colleagues observed that many non-profits they studied moved beyond scaling out because replication alone did not address the systemic problems they cared about. *Scaling up* means changing institutions: policy, legislation, resource flows. It is slower, harder, and often requires partnerships

outside the originating organisation's comfort zone. *Scaling deep* means changing hearts and minds, relationships, and cultural practices. Fraser (2023) has since developed this strand considerably, treating scaling deep as the work of creating the cultural conditions that allow any other scaling to hold.

## Vertical, horizontal, diagonal, functional

From the WHO/ExpandNet (2010) tradition, extended by Milat and colleagues (2016) and Edwards and colleagues (2017). *Vertical* scaling embeds an innovation into institutional systems: policy, governance, financing. *Horizontal* scaling replicates or spreads geographically. *Diagonal* scaling combines the two, on the argument that neither vertical nor horizontal alone is usually enough. *Functional* scaling means adding an innovation to one that has already achieved scale: piggybacking on an existing delivery platform or policy window. Uvin, Jain, and Brown's (2000) earlier four-part formulation (quantitative, functional, organisational, political) predates these and is still worth reading as the origin of the distinction between *more of the same* and *more kinds of thing*.

### KEY FRAMING

Fraser's (2023) argument is that scale deep is routinely under-resourced and misunderstood because its outputs are not quantifiable in conventional terms, but that without it, scaling out and scaling up tend to produce change that does not last.

## Small-to-bigger and big-to-better

Britto and colleagues (2022), writing about parenting programs, offer a useful distinction that cuts across the typologies above. *Small-to-bigger* is the classic route: start with a pilot, demonstrate, expand. *Big-to-better* is the reverse: take an existing at-scale system (for example, a national primary care network) and improve it by integrating or enhancing components, instead of building a new delivery structure alongside it. Big-to-better is under-used in the literature but increasingly relevant where existing systems dominate and new parallel structures are neither affordable nor desirable.

## Spread, replication, diffusion

Already discussed in Section 2.1 (Greenhalgh et al., 2004; Bevan et al., 2024). This deserves a separate direction because "we are going to spread this" is a legitimate choice with different implications than "we are going to scale this."

## Scaling back, un-scaling, de-implementation

The literature on deliberately reducing the reach or scope of an innovation is thin but growing (Norton et al., 2017). An O-DEAR answer of "no, not at this scale" or "scale down first" is a legitimate outcome of the scalability decision. Scaling back is not failure, and the language for it should reflect that.

No organisation scales in only one direction. Most blend. The point of these typologies is to make the blend deliberate. An organisation can name that it is scaling out through partnerships, scaling up through a policy strand, and holding the cultural work through a community of practice, instead of assuming one direction covers all three.

**A FULLER TYPOLOGY: EDWARDS, RILEY & WILLIS (2017)**

Edwards, Riley, and Willis (2017) offer one of the more complete frames for thinking about scaling in practice. Five dimensions sit alongside one another, with each generating its own set of choices:

DIMENSION	SUB-GROUPS	EXAMPLES
<b>Object of scaling-up</b>	Discrete innovations; complex innovations	A vaccine or app; an integrated service delivery model
<b>Adaptations</b>	Incremental; continuum-based; disruptive	Updating educational materials; adding a new population group; rethinking the causal pathway
<b>Directions</b>	Vertical; horizontal; diagonal; functional	State policy change; geographic replication; policy-plus-replication; piggy-backing
<b>Pathways</b>	Linear; non-linear	Replicating a screening program; task-shifting rules that open new delivery forms
<b>Influence</b>	Localised; systemic	Hospital-based champions; a social norm shift

Edwards and colleagues (2017) also identified six recurring challenges in scaling efforts that are worth holding in view when planning:

- 1 **Underestimating resources required.** Scaling typically costs more than pilot-stage estimates suggest, particularly when reaching later-adopting populations.
- 2 **Political and policy naivety.** Scaling almost always brushes against regulatory, policy, or funding arrangements. Those who have not mapped the political economy are caught out.
- 3 **Lack of attention to sustainability in early testing.** The pilot that was not designed with scaling in mind is almost always harder to scale than the one that was.
- 4 **Over-emphasis on vertical or horizontal spread.** Vertical alone institutionalises without ground-level uptake; horizontal alone replicates without institutional anchoring.
- 5 **Uneven attention to time.** The scaling literature tends to focus on spatial reach and pays less attention to how long scaling takes, how long effects persist, and how a scaled innovation is refreshed.
- 6 **Inattention to demand.** Supply-side thinking dominates. Demand-side thinking (communications, marketing, community readiness) is often an afterthought.

## 2.5 Optimal scale and trade-offs

Scaling almost always involves trade-offs, and there is rarely a single right answer about how far to go. McLean and Gargani (2019) call the principle **Optimal Scale**, and frame it as a correction to the assumption that bigger is always better. Their argument is that innovators should optimise. The aim is to find the scale at which the benefits most justify the costs, not to maximise.

*Bigger isn't always better. The question isn't how far this can go. It's how far it should.*

— AFTER MCLEAN & GARGANI, 2019; FRASER, 2023

The trade-offs vary by innovation. Fidelity tends to erode with reach: a program delivered across more sites, by more practitioners, will typically hold its original form less tightly, and the question is whether the benefits of reach outweigh the cost of erosion (Stirman et al., 2019; Carroll et al., 2007; Hasson, 2010). Resource demands rarely scale linearly: reaching the hardest-to-reach populations is almost always more expensive per participant than reaching the easier ones (Edwards et al., 2017). And impact is not guaranteed to scale with the intervention. Some effects depend on conditions that cannot be reproduced at scale.

There is a second kind of trade-off that is harder to quantify. Hawe, Shiell, and Riley (2009) argue that interventions should be understood as *events in systems*, not as packages that travel. Scaling a system-disturbing event involves changing the system it arrives in. If the system absorbs the intervention without changing, the intervention has probably been watered down. If the system resists, scaling fails. In practice, most scaling efforts do both in different places, and Optimal Scale is the working judgement about which mix is acceptable.

## 2.6 From evidence to tool: the O-DEAR questions

TOOL 01 · SCALABILITY DECISION

### O-DEAR

*Five questions to help guide the decision between pilot and scale.*

O-DEAR is a set of five questions that sit at the decision point between pilot and scale. Each question maps onto a cluster of evidence discussed above. In workshop settings I use O-DEAR as a conversational prompt, not a scoring tool. The aim is a structured judgement, not a checklist.



FIGURE 4. THE O-DEAR QUESTIONS

**O. ORGANISATIONAL CLARITY**

How does scaling this align with the organisation's mission, strategic priorities, and capacity? Are there limits to growth that would be hit at a particular scale, and what would they cost to manage? This question draws directly on the Justification principle (McLean & Gargani, 2019), Kohlhoff and colleagues' 2026 findings on the importance of a considered business mindset, and Seelos and Mair's (2017) argument that organisations should separate the work of innovating from the work of scaling and sequence them carefully.

**D. DESCRIBABLE**

Can the mechanism that produces the change be described in plain language, and understood by a range of stakeholders? Describable is the O-DEAR translation of the Scaling Focus. If the core cannot be named, the scaling decision has nothing to act on (Nesta, 2014; Barker et al., 2016; Edwards et al., 2017).

**E. EFFECTIVENESS**

Will scaling address the need? Does the innovation produce the intended outcomes in the piloted setting, and what is the quality of the evidence? This maps onto technical justification (McLean & Gargani, 2019; Bradach, 2003), and onto the empirical work on scalability assessment (Zamboni et al., 2019; Ben Charif et al., 2022; Milat et al., 2020). It is deliberately a question about the evidence, not about confidence or intuition.

## A. ADAPTABILITY

What components are fixed and what can flex? Where is the fidelity bound, and where is adaptation both acceptable and expected? This is the O-DEAR entry point into the fidelity-adaptation tension that Part Three develops (Chambers, Glasgow, & Stange, 2013; Stirman et al., 2019; Castro, Barrera, & Martinez, 2004). A pilot that has not answered this question is not ready to scale.

## R. REACH

What is the realistic demand for the innovation, and will the intended population actually engage with it at the proposed scale? This question pulls on the implementation outcomes literature (Proctor et al., 2011), the population-level scaling literature (Milat et al., 2016; Yamey, 2011), and the moral justification strand of McLean and Gargani's framework. Reach that doesn't reach those most affected isn't the scaling that was intended.

## QUESTIONS AND INDICATORS FOR DISCUSSION

Below are the O-DEAR questions and the kinds of information that might inform a discussion. The exercise is interpretive: the right answers depend on context, and the right approach is the one that fits the innovation and the people involved (Ben Charif et al., 2022; Milat et al., 2020; Lee et al., 2023).

O-DEAR	QUESTION	SUGGESTED INDICATORS FOR DISCUSSION
<b>O</b> Organisational clarity	Does scaling this fit our mission, strategic priorities, and capacity? Where are the limits to growth?	Strategic alignment review; capacity audit; sequencing of scaling alongside other organisational work
<b>D</b> Describable	Can the mechanism that produces change be described in plain language? Can a stakeholder repeat it back?	Documented Scaling Focus; articulated theory of change; named minimum specifications
<b>E</b> Effectiveness	Does the innovation produce the intended outcomes? What is the quality of the pilot evidence?	Pilot outcomes data; effect sizes; mechanism-to-outcome linkage; evidence quality assessment
<b>A</b> Adaptability	What components are fixed and what can flex? Where is adaptation expected, and who decides?	Named minimum specifications; FRAME-consistent adaptation log; documented decision points
<b>R</b> Reach	Will the intended population engage at the proposed scale? What is realistic demand, and who might be missed?	Reach and participation figures; demand assessment; community readiness; equity check

*Pathways are rarely linear. Scaling happens in a system that pushes back.*

*Scaling Pathways · Scaling Canvas*

# 03

---

HOW WILL WE SCALE?

**Scaling**

*Part One ended at the point of decision. Part Two picks up once an organisation has decided to scale and is working out how.*

The planning task has three recurring features. First, scaling happens in a complex system, so plans have to allow for things not going as expected. Second, there is no single pathway that suits every innovation: the pathway has to be chosen deliberately, usually with trade-offs. Third, scaling requires a different kind of planning than running a pilot. It pulls in resourcing, business modelling, governance, partnerships, and team design in ways that pilot delivery usually does not.

### 3.1 Scaling in a complex system

The pilot most innovators are working from is usually a *complicated* undertaking: many moving parts, but with a known relationship between cause and effect. Scaling moves the work into a *complex* space, where the relationship between cause and effect is only visible after the fact, and where the system adapts to what the innovator does (Snowden & Boone, 2007).

Three implications follow. First, planning tools built for complicated work (linear project plans, Gantt charts, fixed stage gates) do part of the job but not all of it. They need to be supplemented by tools that handle emergence and feedback. Second, the best scaling strategies are deliberate about what is fixed and what is flexible. Zimmerman's concept of *minimum specifications*, popularised by Westley, Zimmerman, and Patton (2006), is still one of the most useful framings in the literature: name the non-negotiable elements of the innovation, and let everything else be local. Third, evaluation has to keep pace with what is happening, not arrive a year late. Patton's *developmental evaluation* (Patton, 2011; Patton, McKegg, & Wehipeihana, 2015) is the clearest articulation of what evaluation looks like in this space.

#### COMPLEX, NOT COMPLICATED

A useful way of understanding the difference between contexts. Pilots are usually *complicated*. Scaling is *complex*, because of the system the innovation lands in, and that complexity calls for a different mode of thinking.

Hawe, Shiell, and Riley's (2009) formulation applies here too: interventions are events in systems, and the system will not leave them alone. Designing a scaling plan that treats the system as inert produces disappointing outcomes. Designing a plan that *expects* the system to push back produces plans with more slack in them, more feedback loops, and more opportunities to adapt.

## 3.2 Scaling Pathways

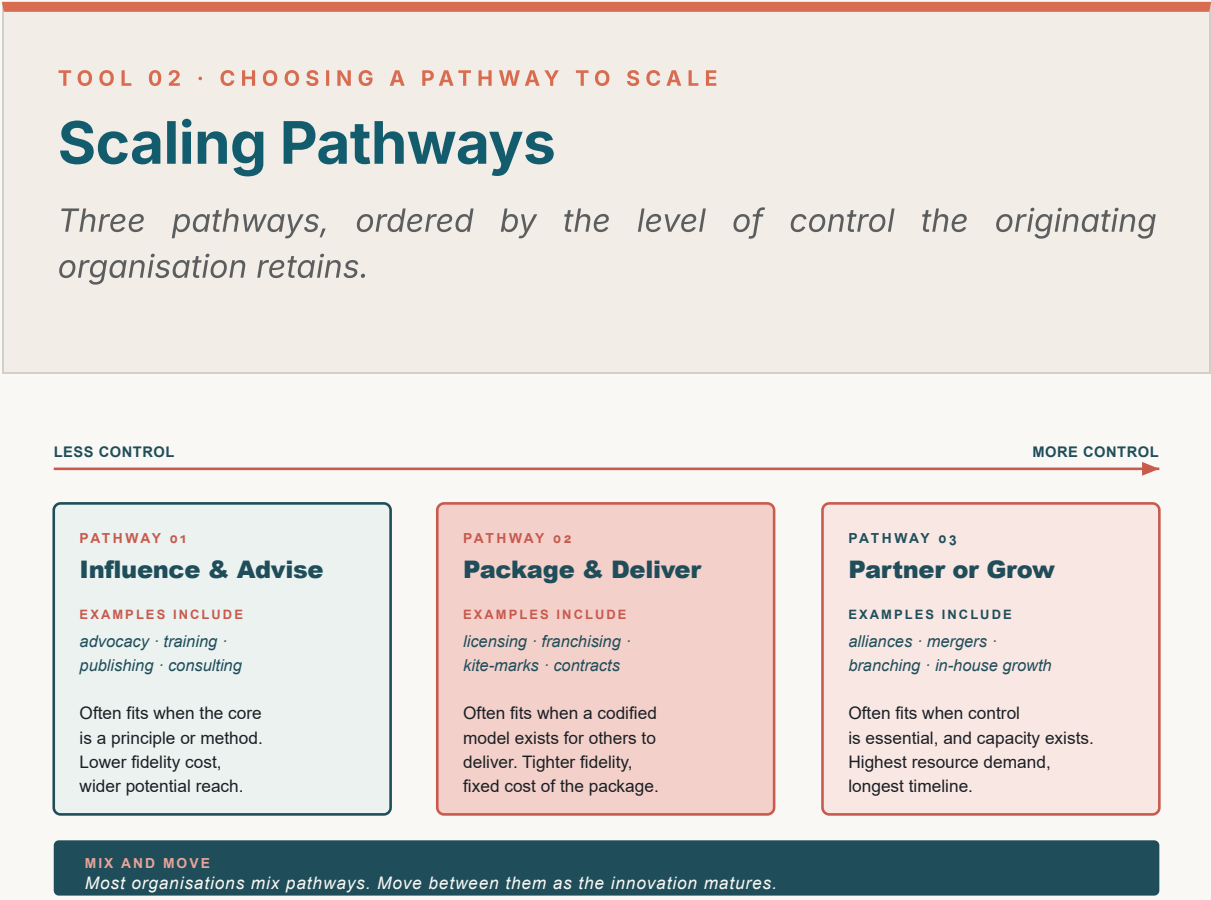


FIGURE 5. THREE PATHWAYS FOR PLANNING TO SCALE, BY LEVEL OF CONTROL. TREAT AS A PLANNING FRAME, NOT A FIXED TAXONOMY.

These three pathways sit comfortably alongside the published typologies. Cooley and Linn's (2014) grouping of *expansion*, *replication*, and *collaboration* maps onto the same logic, distinguished by the degree to which the originating organisation keeps hold of implementation. Nesta's (2014) four routes are the direct source of the FPC three, with two of Nesta's categories collapsed into a single "Partner or Grow" pathway. Dees, Anderson, and Wei-Skillern (2004) set the same logic out as *dissemination*, *affiliation*, and *branching*. Indig and colleagues' (2018) Australian study of actual scaling pathways in public health identified four empirical patterns, reinforcing that the path to scale is rarely linear and that pilot-to-scale stages are frequently skipped.

## NESTA'S FOUR ROUTES IN FULL

SCALING ROUTE	MODELS & APPROACHES	ACTIVITIES	TRADE-OFFS
<b>Influence and advise</b>	Campaigning; consultancy; training	Public speaking; publishing; engaging policy-makers; advising	Low control; wide reach. Best for principles and methodologies.
<b>Build a delivery network</b>	Federations; communities of practice; licensing; franchising	Codifying processes; training; quality assurance; movement-building	Shared purpose with codification; deliberate quality arrangements required.
<b>Form strategic partnerships</b>	Strategic alliances; mainstreaming; joint ventures; M&A	Brokering partnerships for step-change in scale; transferring knowledge	"Bees and trees" model. Raises brand and values questions.
<b>Grow an organisation to deliver</b>	New branches; growing central delivery capacity	Building staff; raising funds; developing organisational systems	Greatest control and fidelity. Most resource-intensive.

Organisations rarely choose one pathway and stick to it. More often, an innovation moves across pathways as it matures. An early-stage innovation may start in Influence and Advise (publishing, training) and move into Package and Deliver once a model is codified. It may then build a Partner or Grow strand to reach populations that the packaged model alone cannot. The pathway is a current state, not a permanent identity.

The pathway choice has implications for the business model. Influence and Advise tends to be funded through grants, speaking, and training income. Package and Deliver tends to require investment in producing the package (manuals, training, fidelity supports) and then relies on licensing or subscription income. Partner or Grow relies on contract income, service delivery funding, or a mix. Kohlhoff and colleagues (2026) were specific that early parenting interventions that scaled well had business models that matched the pathway they were trying to travel: mismatched business models are one of the more common ways scaling fails.

### 3.3 Strategic considerations: what makes scaling work

Three factors weigh particularly heavily in determining whether an innovation will scale successfully: clarity over the core of the innovation, sufficient financing and resourcing to deliver and continuously improve, and the quality of networks, partnerships, and collaborations (Nesta, 2014; Edwards et al., 2017; Kohlhoff et al., 2026). Around them sit a wider set of influences that also bear on the outcome.

#### INFLUENCES ON SCALING

INFLUENCE CATEGORY	SUB-INFLUENCES	NOTES
<b>Innovation characteristics</b>	Documented goals; identified Scaling Focus; fit with population; adaptability; effectiveness; fidelity; supply/demand; plain-language explainability; cost modelling	Tying outcomes to a rigid model reduces the capacity to adapt (Chambers et al., 2013).
<b>Outer context</b>	Policy & legislation; system change; funding environment; community ownership; population characteristics; power dynamics	Largely outside the innovator's control but can be influenced (Shelton et al., 2018).
<b>Inner (organisational) context</b>	Culture & climate; leadership; structures; non-financial resourcing; delivery capacity; mission alignment	Under-estimation of late-adopter demands is a recurring failure mode (Damschroder et al., 2022).
<b>Organisational capacity</b>	Champions; workforce attributes; stakeholder support	Champion mix shifts by sector (Stirman et al., 2012).
<b>Processes &amp; interactions</b>	Planning; engagement; shared decision-making; integration of rules; evaluation; training; communications; partnerships	Often the elements most under-specified in written plans.

Schell and colleagues (2013) offer a useful framework for *capacity for sustainability*. Eight domains (environmental support, funding stability, partnerships, organisational capacity, program evaluation, program adaptation, communications, and strategic planning) collectively predict whether an innovation can hold at scale over time. Luke and colleagues' (2014) Program Sustainability Assessment Tool operationalises it.

Below I work through the six strategic considerations that emerged most clearly from Kohlhoff and colleagues' (2026) qualitative study. They map to the influences above and describe what competent scaling looks like in practice.

## Business mindset and financing

Many innovators come to scaling reluctant to think in business terms. Kohlhoff and colleagues (2026) quote program developers describing their work as "the business that didn't want to be a business". Yet every study of successful scaling identifies a realistic business model as a precondition. Osterwalder and Pigneur's (2010) Business Model Canvas remains the most widely used tool: customer segments, value propositions, channels, relationships, revenue streams, key resources, key activities, key partnerships, and cost structure. Wirtz and Daiser (2017) treat the same territory as *business model innovation*.

*Commercialisation isn't the same as profit-motivation. The best scaling teams use business discipline to protect their values, not to replace them.*

— SEELOS & MAIR, 2017

Start-up funds for the scaling effort itself are often underestimated. Kohlhoff and colleagues' participants routinely described using personal funds or small philanthropic grants to pay for the marketing, legal setup, and infrastructure development that scaling demands but that pilot funding does not cover. Bloom and Smith (2010) make the same point through their SCALERS model: revenue generation is a first-order driver of scaling impact, not an afterthought.

## Stakeholder buy-in across levels

Scaling requires support at multiple levels simultaneously: frontline staff, managers and administrators, families and communities, government and policy-makers (Kohlhoff et al., 2026). The implementation science literature has converged on the same point via different routes: CFIR (Damschroder et al., 2009, 2022), EPIS (Aarons, Hurlburt, & Horwitz, 2011; Moullin et al., 2019), and Community-Based Participatory Research (Wallerstein et al., 2017) all foreground multi-level engagement.

## Partnerships and networks

Partnerships extend delivery capacity, bring in expertise (business, legal, marketing, community), provide credibility, and share risk (Edwards et al., 2017; Nesta, 2014; Kohlhoff et al., 2026). Cabaj and Weaver (2016) on Collective Impact 3.0 emphasises community-anchored partnerships. Bate and

### FROM THE EVIDENCE

Kohlhoff et al. (2026) identified practical strategies: early consultation, site visits, flexible training, and financial incentives for clinicians; outreach to decision-makers for managers; genuine listening for communities; policy alignment for government.

colleagues (2014) argue that the most useful partnerships already have a foothold in the context being scaled into. Partnerships run on a slower clock than delivery.

### Team composition and champions

Scaling is a multi-disciplinary activity. Clinical or technical expertise alone is not sufficient. Successful teams include project coordination, business acumen, stakeholder engagement, policy awareness, marketing, and design, plus the core expertise. Fixsen and colleagues (2009, 2017) organise this into *implementation drivers*: competency, organisational, and leadership. Champions matter at all levels: within the originating organisation, within partner sites, within funding and policy, and within communities.

### Time and planning

Nothing in the scaling literature more consistently surprises first-time innovators than how long scaling takes. Kohlhoff and colleagues' (2026) participants described scaling efforts that took many years to reach the point where scaling became feasible. Fixsen and colleagues' (2009) stages of implementation suggest two to four years to move from exploration through to full implementation in a single setting. Funding cycles that assume scaling happens in twelve to eighteen months are funding something else.

### Coordination

McLean and Gargani's (2019) third guiding principle is **Coordination**: a multi-level, collective perspective involving the participation of actors from many standpoints. Coordination is the connective tissue across the other considerations. A strong business model without coordination becomes internally coherent but isolated. Strong partnerships without coordination become a patchwork. The coordinator role is often under-resourced and under-specified in scaling plans.

## 3.4 From evidence to tool: the Scaling Canvas

TOOL 03 · PLANNING TO SCALE

### The Scaling Canvas

*An eleven-component planning canvas. A deliberate hybrid of the Business Model Canvas (Osterwalder & Pigneur, 2010) and a Theory of Change (Vogel, 2012).*

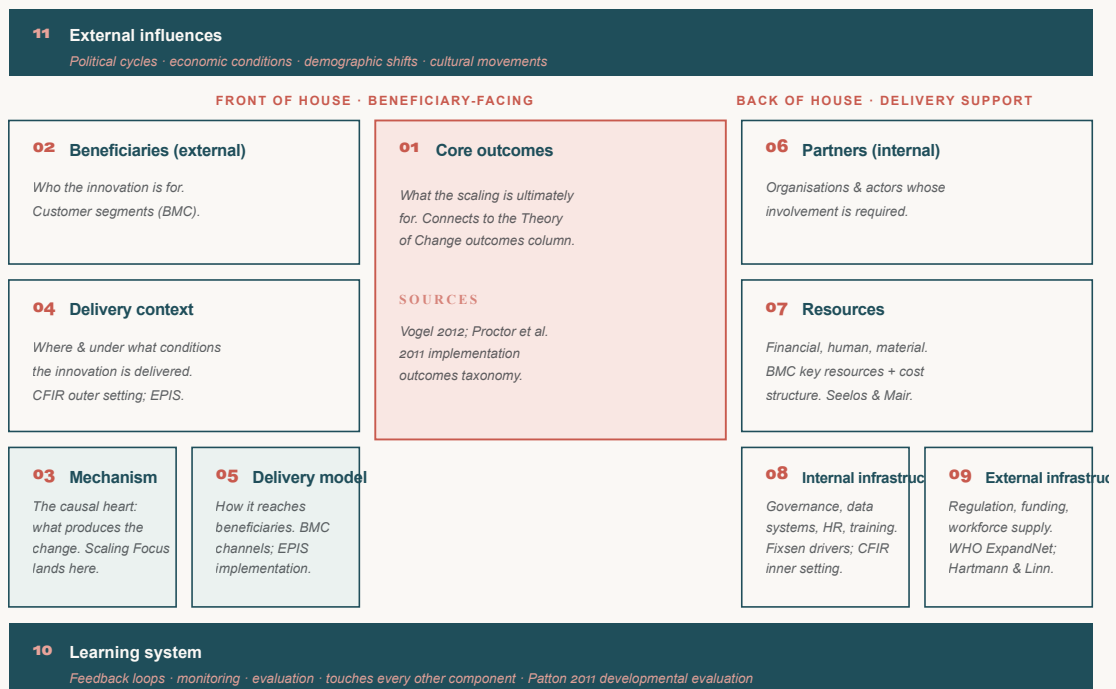


FIGURE 6. THE SCALING CANVAS: ELEVEN COMPONENTS ACROSS FRONT OF HOUSE AND BACK OF HOUSE

The Canvas is laid out with *front of house* on the left (the beneficiary-facing components: what they experience, where it happens, how it reaches them) and *back of house* on the right (the components that support delivery but sit behind the scenes: partners, resources, infrastructure). A top strip holds external influences, and a bottom strip holds the learning system that feeds back into everything else. The central column runs from *core outcomes* at the top (what the scaling is for) down through the mechanism and delivery model that produce those outcomes.

A few things about the Canvas as a whole. It's a one-page articulation of the plan's moving parts, intended to be revisited as the scaling effort unfolds (Westley et al., 2006). Not the plan itself. It assumes complexity: the "learning system" strip at the bottom is built in specifically to reflect that any plan will need to change. And it assumes business thinking and Theory of Change thinking belong together, not in separate documents. Kohlhoff and colleagues (2026) found this combination to be one of the clearer predictors of scaling success.

*Drift is the default. Holding fidelity at scale requires deliberate design.*

TOOL

*Fidelity Checker*

# 04

---

IS THE INNOVATION WORKING AT SCALE?

**Scale**

*Once an innovation is being delivered at scale, the question shifts again.*

The Scalability question is answered (yes, we are scaling). The Scaling plan is in motion. What remains is to check, repeatedly, whether what is happening at scale is still the innovation that was piloted, and whether it is still producing the benefits that justified the decision to scale in the first place.

The checking has three parts. The fidelity-adaptation tension: what holds and what flexes. The implementation drivers: what is carrying the innovation forward at scale. And dynamic evaluation: how the work keeps learning as it goes.

## 4.1 The fidelity-adaptation tension

The tension between fidelity and adaptation is the defining methodological problem of scaled delivery. Too much fidelity, and the innovation cannot adapt to new contexts, so it fails to land. Too much adaptation, and the innovation is no longer the thing that was shown to work, so its effects erode (Castro, Barrera, & Martinez, 2004; Stirman et al., 2012, 2019).

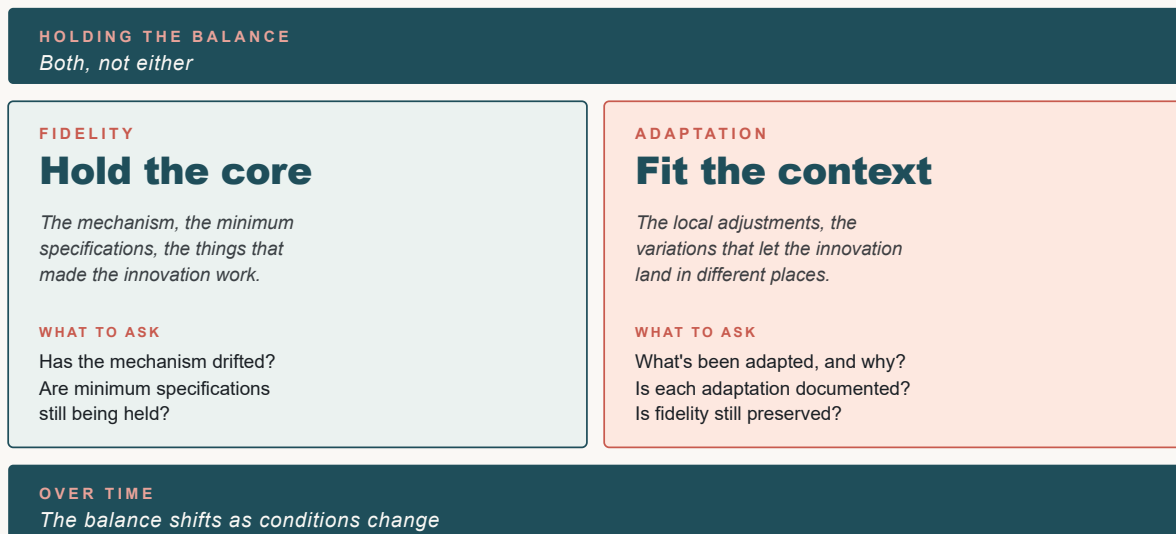


FIGURE 7. FIDELITY AND ADAPTATION AS COMPLEMENTARY, NOT OPPOSITE.

Carroll and colleagues' (2007) conceptual framework for implementation fidelity is still the cleanest starting point. They set out five fidelity dimensions: adherence (content, frequency, duration, coverage), exposure/dose, quality of delivery, participant responsiveness, and program differentiation. Hasson (2010) extended the framework to include context and recruitment as moderators. Stirman, Baumann, and Miller's (2019) FRAME framework gives practitioners a structured way to document adaptations across eight dimensions: when, planned versus reactive, who decided, what was adapted, at what

level, of what type, whether fidelity-consistent, and for what reasons. Miller and colleagues' (2021) FRAME-IS extends the approach to adaptations of *implementation strategies* rather than the intervention itself.

Chambers, Glasgow, and Stange (2013) take the argument a step further with their *Dynamic Sustainability Framework*. Their claim, which has since become the mainstream position, is that the traditional framing (maintaining an intervention in its original form) is inconsistent with how interventions actually behave in real settings. Contexts change, populations shift, evidence accumulates, and rigid maintenance produces quiet decay rather than durable benefit. The dynamic view treats sustainability as a continuous process of managing and supporting the evolution of an intervention within a changing context.

#### FROM THE EVIDENCE

Stirman et al. (2012) found that most interventions experience some drift in real-world conditions. Drift is the default; holding fidelity requires deliberate design.

Three practical implications fall out of this literature. First, the core components (what Zimmerman called the *minimum specifications*) must be named explicitly and kept visible. Second, adaptations should be documented, not hidden; FRAME and FRAME-IS provide ready-made documentation structures. Third, the fidelity-versus-adaptation framing itself is misleading when treated as a binary. Kohlhoff and colleagues (2026) described successful scaling efforts as ones that combined tight core fidelity with deliberate flexibility around the edges.

## 4.2 Implementation drivers at scale

Fixsen and colleagues' work on *implementation drivers* remains the most widely used framework for understanding what carries an innovation forward at scale (Fixsen et al., 2005, 2009, 2017; Bertram, Blase, & Fixsen, 2015). They group the drivers into three classes that are *integrated and compensatory*. Strength in one can partly offset weakness in another, but strength in all three is what sustained scaled implementation requires.

#### FIXSEN'S THREE DRIVER CLASSES

**Competency drivers.** Staff selection, training, coaching, performance assessment.

**Organisational drivers.** Supportive systems, facilitative administration, decision support and data systems.

**Leadership drivers.** Technical leadership for stable contexts, adaptive leadership for changing ones.

The Consolidated Framework for Implementation Research (Damschroder et al., 2009, updated 2022) takes a broader cut at the same territory. Its five domains (innovation, outer setting, inner setting, individuals, and implementation process) overlap with Fixsen's drivers but include more attention to the characteristics of the innovation itself and to the

individuals involved at every level. Nilsen's (2015) taxonomy is the clearest single overview of how these different frameworks relate. Proctor and colleagues' (2011) implementation outcomes taxonomy gives a complementary set of *what to measure*: acceptability, adoption, appropriateness, feasibility, fidelity, implementation cost, penetration, and sustainability.

### 4.3 Dynamic evaluation

*The right question at scale isn't "does it work?" It's "why does it work, and under what conditions?" Continuous re-questioning, not periodic assessment.*

— MCLEAN & GARGANI, 2019

The fourth of McLean and Gargani's (2019) guiding principles is **Dynamic Evaluation**. Their argument is that evaluation of scaled innovations has to move past the question "does it work?" to "why does it work, and under what conditions?". This requires continuous, not periodic, assessment. An evaluation that arrives at the end of the scaling effort is too late to inform it.

The strongest practical expression of this principle is Patton's (2011) *developmental evaluation*. Patton's framing treats the evaluator as part of the innovation team, generating feedback in real time and adapting the evaluation design as the innovation changes. Patton, McKegg, and Wehipeihana's (2015) *Developmental Evaluation Exemplars* collects twelve worked cases, including Indigenous and Māori contexts. Gamble's (2008) *Developmental Evaluation Primer* is a more accessible entry point. Bamberger, Vaessen, and Raimondo's (2016) book offers a practical companion for evaluators in resource-constrained settings.

### 4.4 Sustainability as continued benefit at scale

Sustainability was once treated in the literature as a distinct concept, separate from implementation. The field has since shifted to treat sustainability as one of several implementation outcomes (Proctor et al., 2011; Shelton et al., 2018), and the user-facing question at scale is usually better framed as *is the innovation still producing benefit?* than as *is the innovation sustaining?*

The definitional starting point remains Moore, Mascarenhas, Bain, and Straus's (2017) five-item working definition: *after a defined period of time*, the program continues to be delivered, individual behaviour change is maintained, and both the program and the behaviour change may evolve or adapt while continuing to produce benefits.

Two practical frameworks operationalise this view. Schell and colleagues' (2013) capacity-for-sustainability framework identifies eight domains: environmental support, funding stability, partnerships, organisational capacity, program evaluation, program adaptation, communications, and strategic planning. Luke and colleagues' (2014) Program Sustainability Assessment Tool operationalises this as a 40-item instrument. Maher, Gustafson, and Evans's (2010) NHS Sustainability Model, developed earlier and still in use in health-service improvement, offers a complementary diagnostic.

Two more specific tools deserve attention. Gupta, Thorpe, Bhattacharyya, and Zwarenstein's (2016) *nose-to-tail tool* walks an innovator from initial idea through to long-term uptake, naming at each stage the evidence and strategic decisions that matter. Scheirer and Dearing's (2011) proposed research agenda names seven forms that sustainability takes: continued benefits, continued activities, partnerships, policy or structural change, community attention, diffusion, and unintended consequences.

Sustainability, in short, isn't a separate problem at scale. It's what Part Three is already about. The innovations that survive at scale are the ones that keep adapting, keep being supported, and keep producing the benefits that justified them in the first place.

## 4.5 From evidence to tool: the Fidelity Checker

TOOL 04 · MONITORING AT SCALE

### The Fidelity Checker

*A seven-component review and monitoring tool. Designed to be used regularly (quarterly, six-monthly), not as a one-off compliance check.*

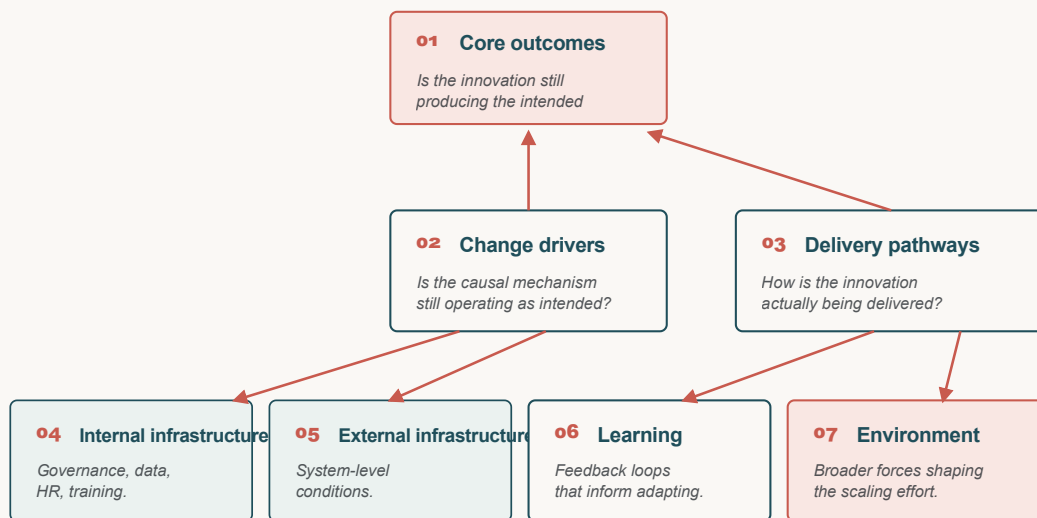


FIGURE 8. THE FIDELITY CHECKER: SEVEN COMPONENTS, WITH ARROWS SHOWING WHAT FEEDS OUTCOMES

The Fidelity Checker is designed to be used regularly, not as a one-off check. It maps back onto the Scaling Canvas so that the monitoring arrangements are directly tied to the planning.

### COMPONENT 1. CORE OUTCOMES

Is the innovation still producing the intended outcomes at scale? This is the O-DEAR Effectiveness question, re-asked at scale. Draws on Proctor and colleagues (2011) and the pilot-to-scale effectiveness literature.

### COMPONENT 2. CHANGE DRIVERS

Is the mechanism (the causal heart of the innovation) still operating as intended? Maps to the Scaling Canvas's *Mechanism* (Canvas component 3). Draws on Carroll et al. (2007), Hasson (2010), and the Scaling Focus framing from Part One. If the mechanism has drifted, outcomes may still look acceptable in the short term but the intervention is no longer what was piloted.

### COMPONENT 3. DELIVERY PATHWAYS

How is the innovation actually being delivered at each scaled site, and how does that compare to the intended delivery model? Maps to Canvas components 4 (Delivery context) and 5 (Delivery model). Draws on Hasson (2010), Dusenbury et al. (2003), and FRAME (Stirman et al., 2019) for structured adaptation documentation.

#### COMPONENT 4. INTERNAL INFRASTRUCTURE

Is the originating organisation's capacity keeping pace? Governance, data systems, HR, training infrastructure. Maps to Canvas component 8. Fixsen et al. (2009) organisational drivers; CFIR inner setting (Damschroder et al., 2022).

#### COMPONENT 5. EXTERNAL INFRASTRUCTURE

Are the system-level conditions still supporting scaled delivery? Funding, regulation, workforce, community infrastructure. Maps to Canvas component 9. Moullin et al. (2019) on EPIS outer-context determinants; WHO/ExpandNet (2010).

#### COMPONENT 6. LEARNING

Is the learning system generating feedback that is actually being used to adapt the innovation? Maps to Canvas component 10. Patton (2011) and the developmental evaluation tradition; Miller et al. (2021) FRAME-IS for strategy-level adaptations.

#### COMPONENT 7. ENVIRONMENTAL FACTORS

What broader forces are affecting the scaling effort, and how is the organisation responding? Maps to Canvas component 11. Hawe et al. (2009) on interventions as events in systems; Snowden & Boone (2007) on complex contexts.

#### TWO DESIGN PRINCIPLES

**A review tool, not a compliance tool.** Using the Fidelity Checker as a checklist produces paperwork and not much else. Using it as a prompt for honest conversation about where drift is occurring, where adaptation has been helpful, and where the system is pushing back. That's what the underlying evidence supports.

**Learning as a first-class component.** The Checker treats learning as one of the seven components, not an afterthought. This reflects the Dynamic Evaluation principle and the consistent finding that the innovations that survive at scale are the ones that keep learning (Chambers et al., 2013; Shelton et al., 2018; Kohlhoff et al., 2026).

*The references and a glossary of scaling terms used through the Field Guide.*

# 05

---

FURTHER READING & GLOSSARY

## **References & Appendix**

## References

### PRIMARY REFERENCES (CITED IN-BODY)

- Aarons, G. A., Hurlburt, M., & Horwitz, S. M. (2011). Advancing a conceptual model of evidence-based practice implementation in public service sectors. *Administration and Policy in Mental Health*, 38(1), 4–23.
- Bamberger, M., Vaessen, J., & Raimondo, E. (2016). *Dealing with complexity in development evaluation: A practical approach*. SAGE.
- Barker, P. M., Reid, A., & Schall, M. W. (2016). A framework for scaling up health interventions: lessons from large-scale improvement initiatives in Africa. *Implementation Science*, 11, 12.
- Bate, P., Robert, G., Fulop, N., Øvretveit, J., & Dixon-Woods, M. (2014). *Perspectives on context*. The Health Foundation.
- Ben Charif, A., Zomahoun, H. T. V., Gogovor, A., et al. (2022). Tools for assessing the scalability of innovations in health: a systematic review. *Health Research Policy and Systems*, 20, 34.
- Bertram, R. M., Blase, K. A., & Fixsen, D. L. (2015). Improving programs and outcomes: Implementation frameworks and organization change. *Research on Social Work Practice*, 25(4), 477–487.
- Bevan, H., Ketley, D., Cawthorne, R., Stavropoulou, C., & Scarbrough, H. (2024). Spreading and scaling innovation and improvement: why the differences matter. *BMJ Innovations*, 10(3), 69–73.
- Bloom, P. N., & Smith, B. R. (2010). Identifying the drivers of social entrepreneurial impact: Theoretical development and an exploratory empirical test of SCALERS. *Journal of Social Entrepreneurship*, 1(1), 126–145.
- Bradach, J. (2003). Going to scale: The challenge of replicating social programs. *Stanford Social Innovation Review*, 1(1), 19–25.
- Britto, P. R., Bradley, R. H., Yoshikawa, H., et al. (2022). The future of parenting programs: Ill uptake and scale. *Parenting*, 22(3), 258–275.
- Cabaj, M., & Weaver, L. (2016). *Collective impact 3.0*. Tamarack Institute.
- Carroll, C., Patterson, M., Wood, S., et al. (2007). A conceptual framework for implementation fidelity. *Implementation Science*, 2, 40.
- Castro, F. G., Barrera, M., Jr., & Martinez, C. R., Jr. (2004). The cultural adaptation of prevention interventions: Resolving tensions between fidelity and fit. *Prevention Science*, 5(1), 41–45.
- Chambers, D. A., Glasgow, R. E., & Stange, K. C. (2013). The dynamic sustainability framework. *Implementation Science*, 8, 117.
- Charif, A. B., Zomahoun, H. T. V., LeBlanc, A., et al. (2017). Effective strategies for scaling up evidence-based practices in primary care: A systematic review. *Implementation Science*, 12, 139.
- Cooley, L., & Linn, J. F. (2014). *Taking innovations to scale: Methods, applications and lessons*. Results for Development Institute.
- Damschroder, L. J., Aron, D. C., Keith, R. E., et al. (2009). Fostering implementation of health services research findings into practice: CFIR. *Implementation Science*, 4, 50.
- Damschroder, L. J., Reardon, C. M., Widerquist, M. A. O., & Lowery, J. (2022). The updated CFIR based on user feedback. *Implementation Science*, 17, 75.

- Dees, J. G., Anderson, B. B., & Wei-Skillern, J.** (2004). Scaling social impact: Strategies for spreading social innovations. *Stanford Social Innovation Review*, 1(4), 24–32.
- Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B.** (2003). A review of research on fidelity of implementation. *Health Education Research*, 18(2), 237–256.
- Edwards, N., Riley, B. L., & Willis, C.** (2017). Scaling-up cancer control innovations. In D. A. Chambers et al. (Eds.), *Optimizing the cancer control continuum*. Oxford University Press.
- Fieldston, E., Terwiesch, C., & Altschuler, S.** (2013). Application of business model innovation to enhance value in health care delivery. *JAMA Pediatrics*, 167(5), 409–411.
- Fixsen, D. L., Blase, K. A., Naoom, S. F., & Wallace, F.** (2009). Core implementation components. *Research on Social Work Practice*, 19(5), 531–540.
- Fixsen, D. L., Blase, K. A., & Fixsen, A. A.** (2017). Scaling effective innovations. *Criminology & Public Policy*, 16(2), 487–499.
- Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F.** (2005). *Implementation research: A synthesis of the literature*. NIRN.
- Fraser, T.** (2023). *The art of scaling deep*. The Systems Sanctuary.
- Gamble, J. A. A.** (2008). *A developmental evaluation primer*. J.W. McConnell Family Foundation.
- Gargani, J., & McLean, R.** (2017). Scaling science. *Stanford Social Innovation Review*, 15(4), 34–39.
- Greenhalgh, T., Robert, G., Macfarlane, F., Bate, P., & Kyriakidou, O.** (2004). Diffusion of innovations in service organizations. *The Milbank Quarterly*, 82(4), 581–629.
- Gupta, A., Thorpe, C., Bhattacharyya, O., & Zwarenstein, M.** (2016). Promoting development and uptake of health innovations: The nose to tail tool. *F1000Research*, 5, 361.
- Hartmann, A., & Linn, J. F.** (2008). *Scaling up: A framework and lessons for development effectiveness*. Brookings Institution.
- Hasson, H.** (2010). Systematic evaluation of implementation fidelity of complex interventions. *Implementation Science*, 5, 67.
- Hawe, P., Shiell, A., & Riley, T.** (2009). Theorising interventions as events in systems. *American Journal of Community Psychology*, 43(3–4), 267–276.
- Indig, D., Lee, K., Grunseit, A., Milat, A., & Bauman, A.** (2018). Pathways for scaling up public health interventions. *BMC Public Health*, 18(1), 68.
- Kohlhoff, J., Donkers, M., & Cibralic, S.** (2026). Scaling early parenting interventions: A qualitative investigation. *Administration and Policy in Mental Health*.
- Lee, K., Crane, M., Grunseit, A., et al.** (2023). Development and application of the Scale-Up Reflection Guide. *International Journal of Environmental Research and Public Health*, 20(11), 6014.
- Lemire, C., Rousseau, M., & Dionne, C.** (2023). A comparison of fidelity implementation frameworks used in early intervention. *American Journal of Evaluation*, 44(3), 383–402.
- Luke, D. A., Calhoun, A., Robichaux, C. B., Elliott, M. B., & Moreland-Russell, S.** (2014). The Program Sustainability Assessment Tool. *Preventing Chronic Disease*, 11, 130184.
- Maher, L., Gustafson, D., & Evans, A.** (2010). *NHS sustainability model and guide*. NHS Institute for Innovation and Improvement.
- McLean, R., & Gargani, J.** (2019). *Scaling impact: Innovation for the public good*. Routledge.
- Milat, A. J., King, L., Bauman, A. E., & Redman, S.** (2013). The concept of scalability. *Health Promotion International*, 28(3), 285–298.

- Milat, A. J., Bauman, A., & Redman, S.** (2015). Narrative review of models and success factors for scaling up. *Implementation Science*, 10, 113.
- Milat, A. J., Newson, R., King, L., et al.** (2016). A guide to scaling up population health interventions. *Public Health Research & Practice*, 26(1), e2611604.
- Milat, A., Lee, K., Conte, K., et al.** (2020). Intervention Scalability Assessment Tool. *Health Research Policy and Systems*, 18, 1.
- Miller, C. J., Barnett, M. L., Baumann, A. A., Gutner, C. A., & Wiltsey Stirman, S.** (2021). The FRAME-IS. *Implementation Science*, 16, 36.
- Moore, J. E., Mascarenhas, A., Bain, J., & Straus, S. E.** (2017). Developing a comprehensive definition of sustainability. *Implementation Science*, 12, 110.
- Moore, M. L., Riddell, D., & Vocisano, D.** (2015). Scaling out, scaling up, scaling deep. *Journal of Corporate Citizenship*, 58, 67–84.
- Moullin, J. C., Dickson, K. S., Stadnick, N. A., Rabin, B., & Aarons, G. A.** (2019). Systematic review of the EPIS framework. *Implementation Science*, 14, 1.
- Movsisyan, A., Arnold, L., Evans, R., et al.** (2019). Adapting evidence-informed complex population health interventions for new contexts. *Implementation Science*, 14, 105.
- Mulgan, G.** (2006). The process of social innovation. *Innovations*, 1(2), 145–162.
- Nesta.** (2014). *Making it big: Strategies for scaling social innovations*. Nesta.
- Nilsen, P.** (2015). Making sense of implementation theories, models and frameworks. *Implementation Science*, 10, 53.
- Norton, W. E., Kennedy, A. E., & Chambers, D. A.** (2017). Studying de-implementation in health. *Implementation Science*, 12, 144.
- Osterwalder, A., & Pigneur, Y.** (2010). *Business model generation*. John Wiley & Sons.
- Patton, M. Q.** (2011). *Developmental evaluation*. Guilford Press.
- Patton, M. Q., McKegg, K., & Wehipeihana, N.** (Eds.). (2015). *Developmental evaluation exemplars*. Guilford Press.
- Proctor, E., Silmere, H., Raghavan, R., et al.** (2011). Outcomes for implementation research. *Administration and Policy in Mental Health*, 38(2), 65–76.
- Riddell, D., & Moore, M. L.** (2015). *Scaling out, scaling up, scaling deep*. J.W. McConnell Family Foundation & Tamarack Institute.
- Scheirer, M. A., & Dearing, J. W.** (2011). An agenda for research on the sustainability of public health programs. *American Journal of Public Health*, 101(11), 2059–2067.
- Schell, S. F., Luke, D. A., Schooley, M. W., et al.** (2013). Public health program capacity for sustainability. *Implementation Science*, 8, 15.
- Seelos, C., & Mair, J.** (2017). *Innovation and scaling for impact*. Stanford University Press.
- Shelton, R. C., Cooper, B. R., & Stirman, S. W.** (2018). The sustainability of evidence-based interventions and practices. *Annual Review of Public Health*, 39, 55–76.
- Snowden, D. J., & Boone, M. E.** (2007). A leader's framework for decision making. *Harvard Business Review*, 85(11), 68–76.
- Stirman, S. W., Kimberly, J., Cook, N., et al.** (2012). The sustainability of new programs and innovations. *Implementation Science*, 7, 17.
- Stirman, S. W., Baumann, A. A., & Miller, C. J.** (2019). The FRAME. *Implementation Science*, 14, 58.
- Taplin, D. H., & Clark, H.** (2012). *Theory of change basics*. ActKnowledge.

- Uvin, P., Jain, P. S., & Brown, L. D.** (2000). Think large and act small. *World Development*, 28(8), 1409–1419.
- Vogel, I.** (2012). *Review of the use of 'theory of change' in international development*. DFID.
- Wallerstein, N., Duran, B., Oetzel, J. G., & Minkler, M.** (2017). *Community-based participatory research for health*. Wiley.
- Westley, F., Antadze, N., Riddell, D. J., Robinson, K., & Geobey, S.** (2014). Five configurations for scaling up social innovation. *The Journal of Applied Behavioral Science*, 50(3), 234–260.
- Westley, F., Zimmerman, B., & Patton, M. Q.** (2006). *Getting to maybe*. Random House Canada.
- Wirtz, B., & Daiser, P.** (2017). Business model innovation: An integrative conceptual framework. *Journal of Business Models*, 5(1), 14–34.
- World Health Organization.** (2010). *Nine steps for developing a scaling-up strategy*. WHO.
- World Health Organization.** (2024). *Designing, implementing, evaluating, and scaling up parenting interventions*. WHO.
- Yamey, G.** (2011). Scaling up global health interventions: A proposed framework for success. *PLoS Medicine*, 8(6), e1001049.
- Zamboni, K., Schellenberg, J., Hanson, C., Betran, A. P., & Dumont, A.** (2019). Assessing scalability of an intervention. *Health Policy and Planning*, 34(7), 544–552.

## Appendix A Glossary of scaling terms

The scaling literature uses a range of overlapping terms. The table below consolidates the most commonly encountered variants. Different authors use the same term with slightly different emphases; clarifying which definition is in play at the start of any scaling conversation is usually worth the time.

TERM	SOURCE	DEFINITION
<b>Scalability</b>	Milat et al. (2013, 2016)	The ability of an intervention shown to be efficacious on a small scale to be expanded under real-world conditions while retaining effectiveness.
<b>Scaling up</b>	Edwards et al. (2017)	Processes to introduce innovations with demonstrated effectiveness through a delivery structure, aiming to improve coverage and equitable access.
<b>Scaling up</b>	Milat et al. (2016)	Deliberate efforts to increase the impact of successfully tested health interventions to benefit more people and foster policy on a lasting basis.
<b>Scaling</b>	Nesta (2014)	Processes through which social innovations grow: standardisation, achieving economies of scale, diffusion, replication, proliferation.
<b>Scaling out</b>	Moore, Riddell & Vocisano (2015)	Impacting greater numbers. Based on the recognition that many good ideas never spread or achieve widespread impact.
<b>Scaling up</b> (social innovation)	Moore, Riddell & Vocisano (2015)	Impacting law and policy. The roots of social problems transcend particular places; innovation must be codified in law, policy, and institutions.
<b>Scaling deep</b>	Moore et al. (2015); Fraser (2023)	Impacting cultural roots. Change rooted in people, relationships, communities, and cultures.
<b>Vertical scaling</b>	Milat et al. (2016); WHO/ExpandNet (2010)	An intervention introduced across a whole system, institutionalised through policy, regulation, and system change.
<b>Horizontal scaling</b>	Milat et al. (2016); WHO/ExpandNet (2010)	An intervention spread across different sites or groups in a phased manner.
<b>Diagonal scaling</b>	Edwards et al. (2017)	A combination of vertical and horizontal, emphasising that neither alone is usually enough.
<b>Functional scale-up</b>	WHO (2010)	Piggy-backing an innovation onto one that has already achieved widespread acceptance.
<b>Small-to-bigger</b>	Britto et al. (2022)	The classic scaling route: starting with a pilot, demonstrating effect, expanding through a delivery structure.
<b>Big-to-better</b>	Britto et al. (2022)	Taking an existing at-scale system and improving it by integrating or enhancing components.

<b>Replication</b>	Dees et al. (2004)	Reproduction of a program's results across new settings, without requiring every feature of the original to be reproduced.
<b>Sustainability</b>	Moore et al. (2017)	After a defined period, the intervention continues to be delivered and individual behaviour change is maintained, both able to evolve while continuing to produce benefits.
<b>Dynamic sustainability</b>	Chambers et al. (2013)	Managing and supporting the evolution of an intervention within a changing context, rather than maintaining it in original form.

# The FPC Scaling Field Guide

*Four free tools for organisations deciding whether, how, and when to scale an innovation.*

## THE TOOLS

**O-DEAR** · Scalability decision

**Scaling Pathways** · How to scale

**Scaling Canvas** · Planning canvas

**Fidelity Checker** · Monitoring at scale

## CONNECT

🌐 WEB

[www.fpconsulting.com.au](http://www.fpconsulting.com.au)

🌐 LINKEDIN

[linkedin.com/company/first-person-consulting-pty-ltd](https://linkedin.com/company/first-person-consulting-pty-ltd)

📄 FIELD GUIDE

[fpconsulting.com.au/scaling-field-guide.html](http://fpconsulting.com.au/scaling-field-guide.html)

✉️ EMAIL

[admin@fpconsulting.com.au](mailto:admin@fpconsulting.com.au)



CREATIVE COMMONS  
BY-NC-SA 4.0