

Evaluating Prototypes

Prototyping is a fast, low-cost, low-risk, learning-rich approach to surface and test promising responses to tough challenges. The approach originated in the technical fields of engineering, software design, and manufacturing, and has since spread to the arena of social innovation where diverse groups are experimenting with new ways to tackle vexing social and environmental challenges.

Prototyping precedes, rather than replaces, conventional pilot projects. Unlike pilots, where a promising intervention is “fixed” for a longer period of time and assessed through thorough evaluative techniques, prototyping can be used to quickly and inexpensively develop and test ideas that may warrant eventual pilot testing. In some unusual cases, the results of prototype tests may be sufficiently robust that innovators can skip a full-fledged pilot study altogether.

Social innovators are now using prototyping processes in an impressive variety of settings:

- Different types of interventions (e.g., technology, business processes, policies)
- Diverse domains (e.g., criminal justice, environmental sustainability)
- Unique processes (e.g., human-centered design, quality improvement, replicating ‘best practices’)

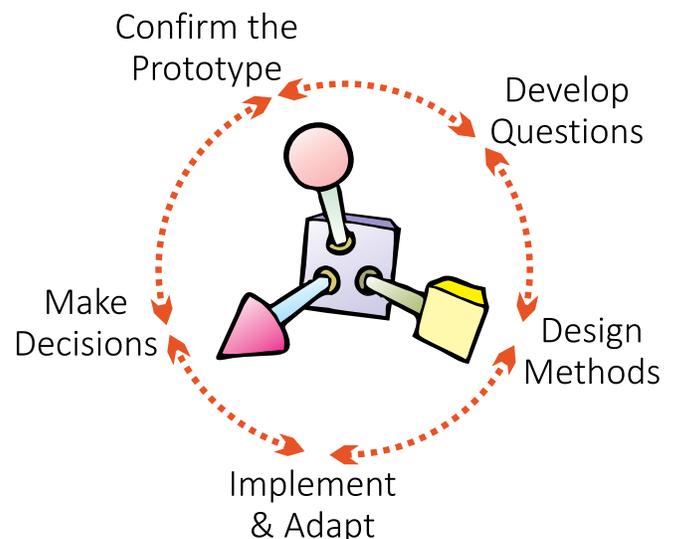
While the art and science of prototyping is growing, comparatively little attention has been paid to how to evaluate the prototypes that emerge from the process.

This document describes 5 steps to the evaluation of prototypes:

1. Confirm the Prototype
2. Develop Questions
3. Design Methods
4. Implement & Adapt
5. Make Decisions

These five steps are iterative. It is a framework, not a recipe, that can guide social innovators and evaluators in designing an evaluation for their prototype. It is also a work in progress: it will be constantly updated based on the experiences of The Natural Step organization in developing its Sustainability Transition Lab model over the course of 2016 and 2017.

Evaluating Prototypes – An Iterative Process



Step 1: Confirm the Prototype



Members of a management team use lego to create multiple rough prototypes of a new business process

The first task in evaluating a prototype is settling on the prototype you want to evaluate. This is trickier than you might think and requires social innovators and evaluators to answer three basic questions:

Do you actually have a prototype?

A prototype is a physical or experiential representation of an idea and how it might play out in the real world. A description of it on a white board, ipad, clip chart or memo will not do. Even very rough prototypes that are tangible enough to evoke feedback from would-be users is sufficient to kickstart the evaluation process.



The urbanistas at the Urban Block Foundation create a temporary “European” block of trees, cafes and bike lanes in an industrial neighborhood in Dallas to engage residents in a conversation about urban design.

What type of prototype is this?

There are a lot of different types of prototype, each with a different purpose and features. As the table on page 4 shows, these include rapid versus slow prototypes, exploratory versus developmental prototypes, one-off or throw-away prototypes versus evolutionary and incremental ones.

This will influence how extensive an evaluation you will carry out. For example, the creators of a lego representation of a new business process can benefit from rapid and light-weight feedback from their peers in a room. By contrast, the creators of an elaborate mock-up of a new pedestrian-friendly urban neighborhood will require something more planned and sophisticated.



An independent inventor creates a small-scale wind turbine to demonstrate how home owners can create their own micro-generator.

What parts of the prototype do you want to test?

Social innovators should be clear if they want to test the whole prototype or just parts of it. The creators of a new small-scale wind turbine may want to get feedback on the entire process of producing, marketing, distributing and servicing the product. Alternatively, they may want to focus just how it might be set up and maintained by the consumers that buy it.

Social innovators and evaluators should not proceed to the next phase of the evaluation process until they can agree on the answers to these questions.

TABLE 1: THE CONTINUUM OF EXPERIMENTS

IDEA	RAPID PROTOTYPE	FIELD PROTOTYPE	PILOT	DECISION
<i>Surfacing new ideas</i>	<i>Making ideas tangible</i>	<i>Testing the “manifestations” of an idea in the field</i>	<i>A full, robust, longer term test of the idea</i>	<i>A decision to adopt, scale, or let go of an idea</i>

- | | | | | |
|--|---|---|--|---|
| <ul style="list-style-type: none"> • Concept Paper • Presentations | <ul style="list-style-type: none"> • Story Boards • Client walk through • Role playing • Lego or paper • Simulations | <ul style="list-style-type: none"> • Simulations • Working samples of all or parts of new model | <ul style="list-style-type: none"> • Pilot projects • Demonstration projects • Trials | <ul style="list-style-type: none"> • (In)formal adoption of policies, regulation, technologies, practices and behaviors that survive the earlier phases. |
|--|---|---|--|---|

Example from the Better Block Foundation’s Work on Urbanizing Forgotten Dallas Neighborhoods (www.betterblock.org)

The group surfaces ideas on how to make Dallas neighborhood more vibrant.

The group sketches out what ideas might look like through vignettes and diagrams.

The group creates a mock up block with trees, bike lanes, cross walks, patios and stories in forgotten neighborhood.

The group elects to carry out more in-depth experiments on key prototypes.

City Council change a score of “anti-street life” regulations, the community introduced bike path, and local entrepreneur opened business based on prototype mock up.



TABLE 2: DIFFERENT TYPES OF PROTOTYPING

TYPE	DESCRIPTION	EXAMPLE
Rapid	A process to help innovators make ideas tangible, learn by doing, and quickly test whether they hold promise for further development.	A new group of professionals and clients of the family justice system develop and sketch out a half-dozen different ways that the legal services can be made more affordable for low-income families.
Slow	A process to develop and test a new idea or model that requires changes in the capacity or culture of the innovating organization in order to succeed.	An oil and gas company diversifying into other forms of energy technology (e.g., solar, thermal, wind) must also build new skills, networks and knowledge in order to be successful in those industries.
Probing (aka Throwaway)	Developing and testing an idea primarily to help groups learn more about the complex system in which they are operating and/or trying to change. Once the experiment is over, the prototype is ‘thrown away’ in the sense it may not be further developed.	An environmental group develops a clean technology prototype to surface the kind of regulatory barriers and opportunities involved with introducing new technology to the market.
Exploratory (aka Proof of Concept Testing)	An early form of prototype, usually a visualization of a concept or idea, tested to determine whether it warrants further development and testing in the field.	A group of local foundations interested in testing a collaborative granting process create a story board that describes each step in detail. It then tests the feasibility of the process with a role-playing exercise with prospective grantees.
Developmental (aka Live)	An applied or field test of an idea, either the entire model or a specific component, in the real world.	A local transit authority expanding its light rail transit line tests the compatibility of its new signalling system with the signalling systems of the older line using several types of relational software.
Incremental	A building block approach where an innovator adds and tests components of a more complex model or solution incrementally over time.	A construction company interested in building “net zero” ready homes develops, tests and refines different elements of the housing unit over time (e.g., water, electrical, heating). It adds and integrates each element as it emerges until the house is considered “net zero” ready.
Evolutionary	Built in an iterative and incremental manner in an effort to more fully develop and refine a model or intervention. May result in prototype 1.0, 1.1, 1.2, 2.0, 3.0, etc.	A cell phone company develops successive versions of a new cell phone to test customers’ responses to an ever more sophisticated set of features.

Step 2: Develop Questions

Questions form the foundation of an evaluation of a prototype. Questions determine which methods and indicators will be used to test the prototypes. The table below contains some typical questions asked of prototypes, with a list of illustrative responses generated by a series of “potential users and supporters” of a hypothetical wind turbine, developed by social innovators interested in options for locally-owned micro-generation.

	Question	Example
The Prototype	To what extent is this prototype likely to be effective in achieving what we want to achieve?	People agree that the only way that this thing would work is if we put it 20 metres tall or higher.
	To what extent is this prototype likely to be feasible in the real world?	These devices are easy to build and they seem as easy to set up as satellite dishes.
	To what extent is this prototype likely to be viable in the current context (economic, political, social, etc.)?	There appears to be very little in the way of municipal regulation of this type of technology: it’s a pretty grey area, which may or may not work in our favour.
	To what extent is this prototype likely to be supported by key stakeholders?	We did not test this in any meaningful way – we should do so in our next iteration and round of testing.
	To what extent is this prototype scalable for bigger impact?	The executive director of a large environmental organization estimates there is potential for 50% market penetration in the city: let’s consider doing a market analysis.
The Challenge	What are we learning about the challenge we are trying to address?	A lot of the people we talked to said that they were looking for ways for homeowners to get involved in micro-generation in general, not just this windmill. They gave us three new ideas for micro-generation.
	What are we learning about the broader context and/or systems in which our challenge and opportunity are embedded?	Local community leagues have a lot of influence with City Hall. We should consider engaging them if we choose to develop the next prototype.
The Team	What did we learn about the capacity of our innovation team?	Boy, our team is more risk adverse than I thought: we were really nervous about developing this idea.
	What has changed in the working relationship amongst our team members?	Everyone on the team agreed that we worked well together and that we should consider doing some new prototypes together.

It is important for social innovators to develop clear questions. “I just want to know what people think!” or “What are the chances this will work?” are a good start. However, they require more crafting in order to get the kinds of response required to truly test the prototype, and provide guidance in the development of more sophisticated methods for assessment.



The National Science Foundation User-Friendly Handbook for Mixed Method Evaluation provides a template for prioritizing and eliminating evaluation questions. See http://betterevaluation.org/resources/guide/design-process_mixed_method_evaluations

Step 3: Design Methods

There is no cookie cutter set of indicators or methods to evaluate prototypes. The evaluation design for the urban charrette, new business model, and client services highlighted on page 2 of the document each require a customized set of tools and indicators.

There are, however, principles to guide social innovators and evaluators when crafting an evaluation design:

- 1. Relevant.** The methods must be designed to answer the key questions of the prototype team.
- 2. Credible.** The evaluation design should be considered legitimate by the people using the evaluation feedback. (For example, social innovators may feel that feedback from engineers on a new carbon-capture technology is more appropriate than the opinions of municipal administrators.)
- 3. Appropriate burden of Proof.** The methods should generate data that meets an appropriate burden of proof. For example, the burden of proof in an exploratory prototype is much lower than in a pilot project where the stakes and risks are higher. (See next page for more details.)
- 4. Quality.** The evaluation design should meet the basic quality standards of the Canadian Evaluation Society.
- 5. Timely.** The results of the assessment should be made available in 'real time,' that is, provided in such a way that social innovators can incorporate the feedback into their decisions about the next steps for the prototype.



Social innovators might be able to draw on these principles to craft evaluations on their own – particularly for evaluating rapid prototypes. For prototypes that require a more sophisticated design, they may require the support of an evaluator.

Whether simple or sophisticated, all evaluation design for prototypes has to answer the following four questions:

- What are the preferred data sources and method for each evaluation question?
- When do the innovators require the feedback?
- What things should be considered in making the evaluation useful?
- Who is the for coordinating and managing the evaluation?

An Escalating Burden of Proof

Rapid Prototypes

Evaluating rapid prototypes should employ quick and light-weight feedback methods, reflecting the low risk stage and stakes at this stage of the process, and focus primarily on getting rapid feedback from would-be users or participants about a new idea.

Example Methods

1. **Appreciative Inquiry:** a way of providing real-time feedback, organized around positively framed questions: e.g., I like this part of the prototype: have you thought about [concern, question, critique]?
2. **Ritual Assent-Dissent:** a structured process by which two teams provide rounds of positive and then negative feedback.
3. **Red Team-Blue Team:** a role-playing game used in military and security fields in which two groups are organized to “defend or attack” an idea, model or approach.

Field Prototypes

As the investment, risks and stakes of the innovation increase, so does the need to complement user feedback with more structured research on key questions. While direct user feedback is still central in prototypes, the questions and methods tend to be more robust and intensive.

Example Studies

1. **Feasibility Analysis:** A study to test select the technical, operational, economic and political feasibility of a new idea, venture or model.
2. **Process Evaluation:** An exploration of the different ways of designing and delivering a new model, surfacing challenges and how they might be resolved.
3. **Outcome Evaluation:** A focus on testing the likely outcomes of the model in the real world.



Step 4: Implement & Adapt

Implementing an evaluation design is rarely as simple as ‘plan the work, work the plan.’ In fact, there are three major reasons that social innovators and evaluators may need to adapt the design and implementation of their evaluation plan:

- The original design is insufficient: A group realizes quickly that adjustments are required to provide good answers to the question. (“We thought we could test this with a focus group of municipal administrators, but it turns out that we need to talk to the planning department as well, and it’s best done with one-on-one interviews.”)
- Deeper questions emerge: The very act of testing a prototype will result in new questions about that prototype. (“We realize that there are a lot more regulatory barriers that we first anticipated. How are we going to get a handle on them all?”)



- The prototype evolves: Social innovators are always upgrading their prototypes and even little adjustments may open up the need for new methods and data. (“We had not considered the option of creating neighborhood-based cooperatives for this technology. How would that sell?”)

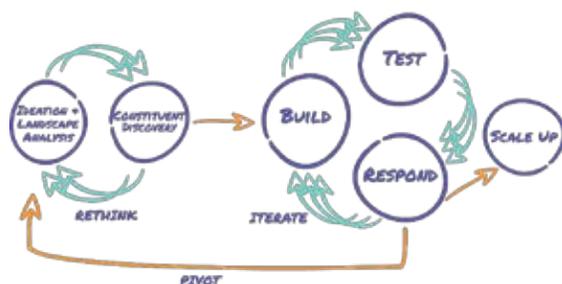
As a general rule, the evaluation for any prototype is a work-in-progress whose design co-evolves as quickly or as slowly as the questions and data expectations of the social innovators.

While rapid prototypes are usually developed and evaluated in a single meeting or session, the work of preparing and assessing field prototypes takes time. Here are three things to draw upon to design your own process.

Lean Start-up Methodology

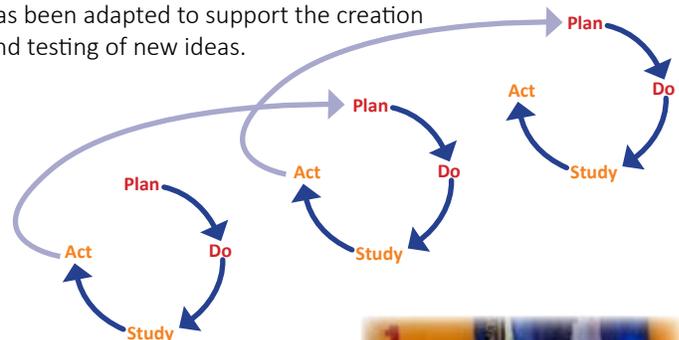
A build-measure-learn process that begins with developing a minimum viable product (MVP) and then testing and learning from it as quickly as possible.

THE LEAN PROCESS FOR THE SOCIAL SECTOR BY ACCELERATE CHANGE



Plan-Do-Study-Act (PDSA) Cycle

This highly structured methodology for experimentation originated in Total Quality Management. Since then it has been adapted to support the creation and testing of new ideas.



Rapid Results Campaign

An idea popularized by the Rapid Results Institute. It encourages social innovators to set micro “stretch goals,” e.g., reduce homelessness in a city by 2.5% – in an iterative series of 100 Day Campaigns.



Evaluation Worksheet for Prototypes

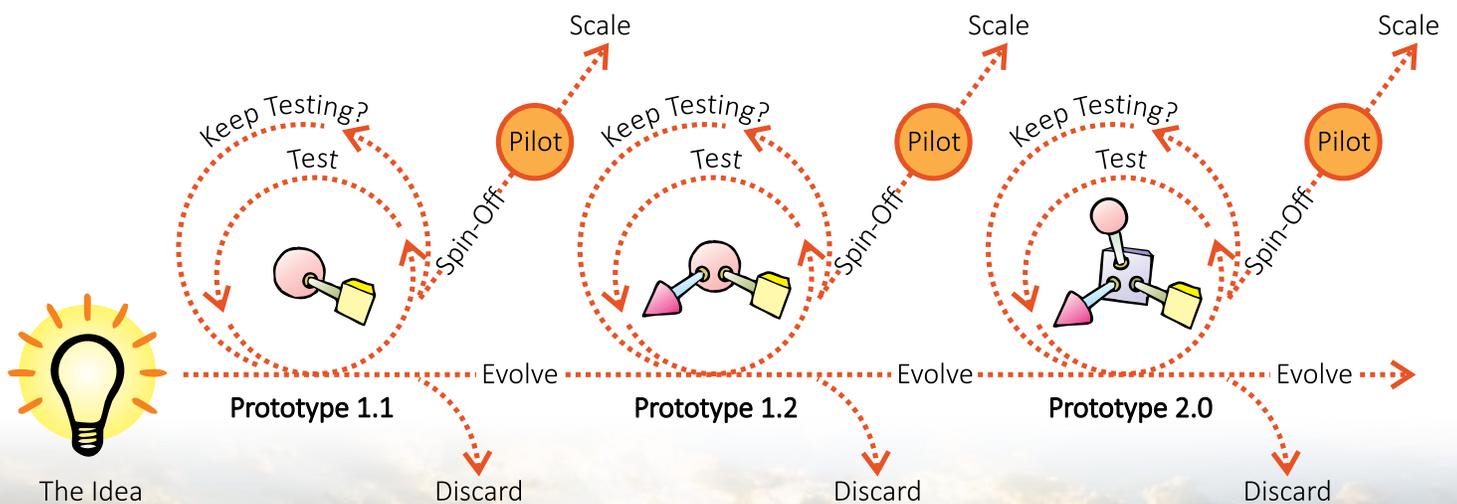
STEP ONE	STEP TWO	STEP THREE			
What part of the prototype is being tested?	Evaluation Questions	Preferred Data Sources/Methods	When Feedback Required	Key Things to Consider to Make Test Useful	Lead/Logistics
A possible “five stages of customer engagement” process for a software company.	Do you think our staff will understand this proposed new business process?	The management team will assume the role of different front-line staff and try to anticipate how each will react to each of the five phases.	Immediately after the first rapid prototypes are produced during the design session.	A facilitator needs to document the feedback from each prototype and give it back to the team ASAP.	Janice, the lead facilitator, will need ipod camera to film the event, with assistant to write up notes that afternoon.
A mock up “European block” in our test neighborhood in Dallas by Better Block Foundation.	Where do our prototypes have friction with municipal policies and regulations?	Have municipal managerial and front line staff from 3 planning departments visit the mock up and write up problem areas on sticky notes, and leave them on site.	During a neighborhood walk through the site facilitated by Foundation staff.	Keep the walkthrough light and fun, allowing participants to offer discrete feedback by email after the session as well.	Christopher, our government relations Director, will lead walk through and Kimberly will transcribe and analyze the notes.
The feasibility and interest of satellite installation experts to install our micro-wind generators on residential roof tops, as part of their expanded set of services.	Does installing micro-generators require extra skills? Under what conditions are installers willing to do it?	Have one on one interviews with ten different installers, from different parts of the city, after a demonstration of installation by our company staff.	By mid-December’s management meeting: this means tests have to be complete in October.	Ensure that the installers in our test group work with a variety of different types of satellites to ensure we get a diversity of expertise.	Janusz will design the whole test project and work with Beatrice – our installation experts – to run the demonstrations.

Step 5: Decisions

The purpose of evaluating a prototype is to help the innovators make a data-informed decision about the future of the prototype.

There are five broad decisions the team can make after reviewing the data and analysis of a prototype evaluation:

- 1. Discard:** The team has decided that the prototype is not worth continued investment.
- 2. Evolve:** Adapt the prototype in some new way:
 - Test the current prototype in a new setting or context.
 - Adapt the current prototype (and evaluation design) based on new learnings.
 - Upgrade the design brief and generate new prototypes.
- 3. Graduate to Pilot:** The test results are sufficiently positive and unambiguous, that the prototype team feels the prototype should be more fully tested through a formal pilot.
- 4. Go to Scale:** The test results are so positive and unambiguous, and the risks so low or manageable, that the prototype team is convinced that it is worth scaling without any further testing.
- 5. Keep Testing:** The results of the evaluation were not strong enough to make a decision at this time. Upgrade the evaluation design and try again.



Resources

Type in 'prototyping' in a google search engine and it turns up over 18,000,000 entries. These five sites are among the most useful of the bunch



<http://www.designkit.org>

The organization that popularized design thinking in general and created a first-rate kit for human-centered design.



<http://diytoolkit.org/media/Prototype-Testing-Plan-Size-A4.pdf>

A little known group, specializing in international development. DIY produced the Prototyping Testing Plan.



Mark Cabaj
Here to There
CONSULTING INC

What We Know So Far is a series of documents that summarize some of the latest thinking or developments in the field of social innovation and community change.

NESTA Making Innovation Flourish

<http://www.nesta.org.uk/>

The world's premier social innovation organization, with a half-dozen tools on prototyping – and a special focus on public services and public policy.



<http://www.servicedesign-tools.org/>

An open-source website with 60 design tools, many of them (e.g., client journeys, user profiles, sketches, simulations) useful for developing and testing prototypes.



<http://www.leanchange.net/blog/>

A webpage with a variety of resources – including a few on evaluation-- organized around the lean start-up methodology, developed in the private sector and adapted for use in social change initiatives.



BetterEvaluation

<http://betterevaluation.org>

The most comprehensive web-based evaluation resource in the world with a step-by-step approach to evaluation design and dozens of tools.

Here to There Consulting Inc. is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported License.