THE Experimentation Field Book

A Step-by-Step Project Guide

Jeanne Liedtka, Elizabeth Chen, Natalie Foley, and David Kester

Columbia Business School Publishing

Columbia University Press
Publishers since 1893
New York Chichester, West Sussex

Library of Congress Cataloging-in-Publication Data

Columbia University Press books are printed on
permanent and durable acid-free paper.
This book is printed on paper with recycled content.
Printed in the United States of America

c 10 9 8 7 6 5 4 3 2 1

Book design by Leigh Ayers, 10/HALF Studios

Contents

Introduction to the Experimentation Process

DOES THE WORD *experimentation* conjure images of test tubes, safety goggles, and possibly a dash of fire or smoke? Our early exposure to experimentation is often in science class as part of the scientific discovery process. Accompanying those beakers and lab coats is the scientific method, a guide to running experiments.

But many people run experiments every day outside of science labs. Long before you were learning about experiments in school, your toddler brain was experimenting as a primary method of learning: *If I stack blocks like this, they don't fall over. If I put my fingers near the door, I can get hurt when it shuts.*

Many organizations use experimentation to explore ideas for new products, services, programs, and strategies. In this field guide, we'll share stories from different **organizations**. Nike wanted to experiment

Organizations in Brief

Nike is one of the largest and most successful athletic apparel companies in the world.

Whiteriver Hospital, located on the Fort Apache Indian Reservation, has both inpatient beds and an emergency room where most of the population's medical needs have to be met.

South Western Railway (SWR), a joint venture between two of the world's leading rail companies, operated some of the busiest train routes in the United Kingdom, with 235 million passenger journeys a year.

The Project Management Institute (PMI) is the premier global project management association, with nearly 500,000 members.

with a shoe subscription service for kids. Whiteriver Hospital wanted to reduce patient wait time in their emergency room. The Project Management Institute (PMI), a professional learning organization, was interested in testing different ways to serve its members. South Western Railway (SWR), one of the UK's biggest commuter railways, hoped to improve passenger experiences. People at each of these organizations had good ideas on paper but wondered if they would succeed in reality. They worried about being wrong in their hunches. So, they **used experimentation to test their hunches *before* they built and launched their new ideas and *before* spending a lot of time and money on them.** Throughout this field book, we will return to these four stories for inspiration and lessons learned as we build your experimentation skills.

Entrepreneurs, teachers, caregivers, marketers, graphic designers, social workers, even people doing their everyday routines, all experiment. Have you ever caught yourself saying, "I wonder if taking this route home is faster? Let me try it and time it." You're experimenting! Or think of a team leader considering, "If we send the agenda in advance, it might make our meetings more effective. Let's try it for a few weeks and see if that's true." Experimenting! This field book focuses on helping you to systematically design and execute your own high-quality experiments on any idea you have.

Experimentation is much talked about in the abstract, but poorly understood in practice. Yet it is the crucial connector between *imagining* an idea and *making it happen* in reality.

Why Experiment?

Here are a few reasons why putting in the time to experiment is better than just building a new idea:

- Experimentation protects you from *overspending* on a solution that won't work for the people you designed it for. By placing small bets and learning at a fast pace, you can learn whether your concept really fulfills your users'—whether patients, passengers, customers, or members—unmet needs and if the idea is really desirable to them.
- It encourages you to test a *portfolio* containing multiple ideas rather than converging prematurely on a single idea. Then it lets your users highlight the one that works best for them.
- Through experiments, you learn how to *scale* your idea effectively and successfully. Testing aspects of the idea at smaller scales reduces the risk of early service delivery friction and is less expensive.
- You'll get to know your *early adopters*, and their feedback will inform future features and help accelerate the development of your new product or service.
- Experimentation is fun and *engaging*! It gets you out of the conference room and into the real world and allows you to invite those who will be part of the implementation of a solution into the testing process, building buy-in and adding energy and momentum to a project.

Experiments exist to collect the data to build an evidence base, to support whether and how to launch new solutions, but they are often underutilized or badly designed. At one extreme, they end up looking like "soft launches" with little appetite for pivoting. In other fields, only elaborate, formal experiments like randomized controlled trials, which often last years rather than days or weeks, count. In this book, we will show you how to take advantage of the often-unexplored territory in between, where well-designed, learning-oriented field experiments can add tremendous learning and risk reduction but take only hours instead of weeks and cost next to nothing.

For most of us, **building an evidence base is the best way to manage the inevitable risk of trying new things in today's dynamic environments**. In fact, skills in the design and execution of experiments are one of the fundamental and critical competencies for success in an increasingly uncertain world where there are known unknowns (things we *know* we don't know) and unknown unknowns (things we *don't* know we don't know). Experimentation helps us to better address the first category and discover the second one. This valuable tool is embedded into many problem-solving approaches (like Lean Startup, Agile Software Development, Design Thinking, Kaizen, and Process Improvement). Experiments also go by many names; you might have heard about a new Google feature that is "in beta" or about organizations running pilots or testing a minimum viable product (MVP). Yes, the terms used for experimentation can be confusing and full of jargon—but the terms we use are less important than being precise about what the experiment seeks to test and how it will accomplish that. Experimentation is a powerful and effective learning tool for individuals or teams who are improving a product or service that exists or creating something new to them, their organization, or the world—all require deliberate learning through action. Though the need for experimentation may seem obvious, many organizations skip it. Why? Because the urge to "just build it" (and hope they will come . . .) is strong.

A hypothesis is the essential starting point for experimentation. Yes, back to science class! A hypothesis is a best guess about what you expect to be true or assume is true: Nike employees could have just assumed parents would want a shoe subscription service and built it, but instead they experimented to test the assumption that it met a real customer need.

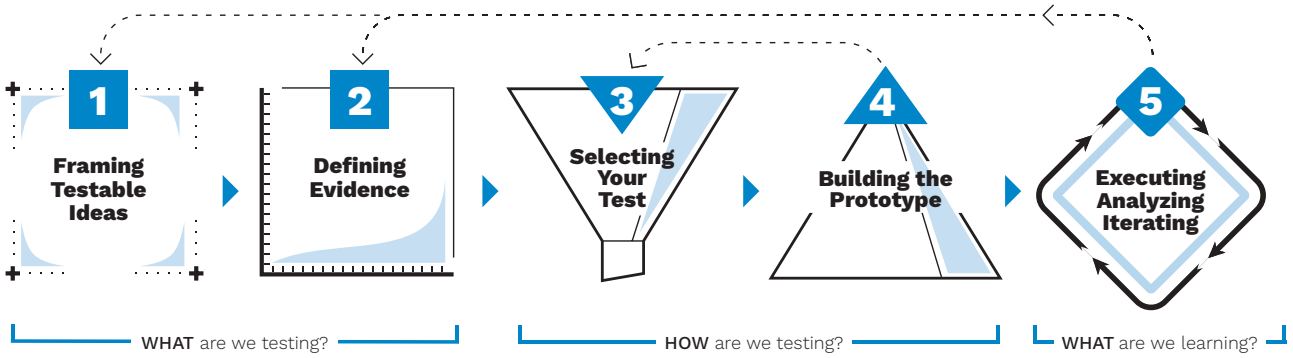
Hypotheses are conjectures that can be true or false. We find out if they're true by collecting data to confirm or disconfirm our hypothesis. For example, you can test if that new route you suspect to be faster truly is by timing it and then comparing that to your old route time. Hypothesis-driven thinking is the anchor of experimentation—and it is fundamentally different than just using existing data to make your decision. Hypothesis-driven thinking flips the usual process that starts with the data you've already got. **Instead, you start with the hypothesis and then identify the data you need to test it rigorously (which is often *not* the data you've already got)**. For your data to be most useful, your data collection must be done with intention. To do that, you need to be hypothesis driven.

Experimentation is not a "one and done" activity. It is a process that evolves over time. You will run multiple tests, which will increase in complexity as your hypothesis is further developed and refined. This de-risks your solution development. Rather than just building the real thing right away, experimenting helps you manage risk (you spend time and money incrementally as risk moves from higher to lower), gather evidence to test your assumptions from a range of actual users (not only from conversations among colleagues or early adopters), and improve your odds of success. The iteration that occurs with experimentation creates a stronger solution by making sure that people really want what you've designed and demonstrating that you can deliver it in a viable way. Experimentation can also be used to understand whether any given problem is even worth solving.

We hope that we've convinced you of the *why* of experimentation and the need to be hypothesis driven. In the rest of this field book, we will focus on the *how* by providing a hands-on, structured process to guide you through the design and execution of your very own experiments. We will outline how to design and run tests to address a variety of challenges—improving an existing solution, fixing something that's broken, making a new product, or launching a new service for your existing customers. To do this, we provide a step-by-step process, using the templates we have created. Together we will explore **five steps**.

Let's look briefly at each step. Steps 1 and 2 focus on the *what* of experimenting. Step 1, framing a testable idea, begins the process. Because we will treat our ideas as hypotheses, we need to consider what makes

The Five Steps



any given idea truly testable. We will explore this as we look at the specifics of your idea and who it serves—information that is needed to design a rigorous test. Next, in Step 2, we define what constitutes evidence—what kind of data will tell us whether our hypothesis (and the assumptions behind it) are true or false? Where will we locate such data? Having specified what we are looking for, we then move onto the question of *how* to conduct our experiment. In Step 3 we need to sort through a variety of options to zero in on the best type of test to collect the desired data. What type of test best suits the particulars of our idea and the evidence we need to gather? Once we have selected our test type, in Step 4 we develop the stimulus we will use to provoke feedback by building the prototype, the simplest one that will do the job. Finally, with the key ingredients in hand—our testable idea (our hypothesis), our test type (our intervention), and our prototype (our stimulus), we move onto the question of what we have *learned* in Step 5, as we execute our test plan, analyze the results it produces, and iterate our way to an improved solution.

But these steps are rarely as linear as our model suggests. Frequently, we loop back to earlier steps. The selection of a test type, Step 3, and prototype format, Step 4, are usually highly interactive, and it is not uncommon to move back and forth in an iterative way to complete them. Sometimes our ideas fail in Step 5 and we loop back to Step 1 to start all over again with a new idea. Or maybe they succeed, so we loop back to Step 2 to design a more sophisticated test with a higher fidelity prototype and different evidence—or to test a different set of assumptions. And, of course, things happen *before* our Step 1—the process of generating the idea, for instance—while the essential activities around scaling an idea in the real world occur *after* our Step 5, when the period of experimentation ends, and implementation begins.

As we move through these five steps and apply them to your idea, we will illustrate them in action with examples from our own experiences and research working with innovators at Nike, Whiteriver Hospital, the Project Management Institute and South Western Railway. Because the ideas you will want to test come in different forms—products, services, processes, software, or a combination of these—our stories encompass this variety. Let's preview these four stories, which we will return to time and time again.

Prototypes and Pyramids

Prototyping has been around a long time! The Great Pyramid at Giza, constructed in 2528 BC, is the oldest of the Seven Wonders of the World and the only one that remains largely intact. Scholars believe that its construction, estimated to take decades—or even centuries—almost certainly relied on an iterative process using low fidelity prototypes, including drawings and schematics. "The wonders of ancient Greece and Egypt required engineering far in advance of their time, and marvels of engineering such as the Pyramids and the Parthenon didn't sprout up out of the ground whole," explains engineering expert David Walsh.*

* The Top Four Ancient Design Prototypes" by David Walsh, 8/19/15, https://www.asme.org/topics-resources/content/top-4-ancient-design-prototypes

Our Four Stories in Detail

Nike *Easykicks*

Context

Nike is one of the largest and most successful athletic apparel companies in the world. Like many consumer-packaged goods firms, they wanted to explore expanding into services that would allow them to diversify in ways that would develop deeper customer relationships.

Challenge

Nike had identified some problems worth addressing: Kids grew fast and wrecked shoes, so parents constantly needed to buy new ones to accommodate their kids' growing feet. Shopping at a store with kids wasn't typically enjoyable. In addition, though some old shoes were donated, many old pairs lay abandoned in musty closets, and Nike, with a commitment to sustainability, had invested in technology to recycle old shoe materials. So, they wanted to explore the idea of a shoe subscription service, *Easykicks*, for young athletes, a population they wanted to get to know and serve better. The new subscription service offered an opportunity to address both the hassle of shoe shopping and the need for recycling.

Process

Nike knew they needed help in testing the business model of services, like *Easykicks*, that had caused challenges for them in the past, so they partnered with Peer Insight, a firm known for acumen in market experimentation. Peer Insight led a process using the steps outlined in this field book to design and run tests on the new subscription service over an 18-month period.

Whiteriver Hospital

Context

Whiteriver Hospital is located on the Fort Apache Indian Reservation which covers more than a million acres and serves a population upwards of 15,000, mostly Native Americans. It falls under the jurisdiction of the U.S. Department of Health and Human Services (H&HS) and has both inpatient beds and an emergency room where most of the population's medical needs have to be met, including prescription refills.

Challenge

The hospital faced a serious situation: close to 25 percent of emergency department (ED) visitors were leaving without being seen, a problem attributed to long wait times. Nonemergency patients consistently got delayed as staff addressed true emergencies, with arrivals sometimes waiting as long as six hours before being seen. When potential patients left the emergency room (which they did at a rate twenty times the U.S. national average), their medical problems worsened. Often, these eventually became true emergencies, and patients needed to be helicoptered off the reservation for more extensive and expensive care. The Whiteriver Performance Improvement Team wanted to explore the idea of adding an electronic kiosk, similar to one at Johns Hopkins Hospital in Baltimore, where a patient electronically signed in upon arrival, and the electronic system informed other parts of the hospital of that patient's potential needs, saving administrative time and speeding up the intake process for patients so that more patients could be seen more quickly in the ED.

Process

H&HS operated the Ignite Accelerator, a program of their IDEA Lab aimed at bringing new innovation approaches to employees across the United States. Ignite offered education, coaching, and a small funding stipend, to boost projects that offered the hope of addressing agency problems. Whiteriver's Performance Improvement Team was invited to test the electronic kiosk idea as part of the Ignite program, using a process similar to that outlined here.

Our Four Stories in Detail continued

South Western Railway

Context

SWR, a joint venture between two of the world's leading rail companies, operates some of the busiest train routes in the United Kingdom, with 235 million passenger journeys a year. SWR faced challenges with industrial relations, major network repairs, staff morale, and the passenger experience.

Challenge

SWR leadership saw an opportunity to improve the passenger experience. Unlike Nike and Whiteriver, SWR did not start with a solution in mind; they wanted to first explore the problems passengers faced in more detail. They commissioned the consultancy David Kester and Associates (DK&A), a leading design firm, to help them find both quick wins and a long-term strategy to improve the SWR customer experience. They captured the challenge as follows: to learn fast from customers, rapidly deliver confidence-building basics at SWR stations, and together shape the future SWR experience.

Process

Using ethnographic research tools, they partnered with SWR staff at three central rail stations, to better understand the problems and needs of their passengers. They synthesized this research, identifying insights and creating journey maps and personas. Based on these, they facilitated a collective brainstorming process that identified multiple ideas. Three concepts emerged as particularly attractive: the *Concierge* concept (aimed at offering a warm welcome and providing the right infrastructure to better support customers in the ticketing process), *Cleaner and Cleaner* (a hygiene awareness and nudge campaign to provide reassurance to the public), and the *Wayfinding* Audit (addressing poor signage and visual clutter with smart customer-led tools). DK&A then embarked on an experimentation phase, using a process similar to the Steps outlined here.

Project Management Institute

Context

PMI is the premier global project management association, with nearly 500,000 members. Like many membership associations, they struggle with how to evolve and continue to add value to their members outside of their core product, the Project Management Professional (PMP) certification.

Challenge

PMI had already done extensive research to better understand the problems and needs of their members, and had identified four high-level ideas they thought had significant potential to create value for them:

1. *Snippets*: A microlearning platform that curated quick, bite-sized trainings and resources to fit members' day-to-day learning needs.
2. *Career Navigator*: A self-assessment tool that showed potential career paths based on a member's experience, skills, and interests.
3. *Hive*: A peer-to-peer connection platform that allowed members to get answers to their toughest project management questions by connecting directly to experienced peers.
4. *Spot*: An experiential learning opportunity that matched members with real-world, low-risk opportunities to help them hone needed skills.

Process

Each idea was intriguing but untested, and with an important board meeting approaching, PMI leadership needed to make choices about which ideas to invest in. They partnered with Peer Insight to run a series of experiments over three months on the four offerings to narrow down which they should move forward. From there, Peer Insight ran a series of deeper market experiments over six months on the best performing concepts, again using our five-step process.

Each of these organizations spent time to first understand the problems their customers faced to ensure that they invested in solutions that solved problems that mattered. We will pick up their stories as each enters the experimentation process with particular ideas that they want to test. Though experimentation can be used to identify problems as well as to test solutions, in this Field Guide we will focus on how to test a given solution—like *Easykicks*, the *Electronic Kiosk*, the *Concierge* service or *Career Navigator*—to see whether it does, in fact, meet an identified need.

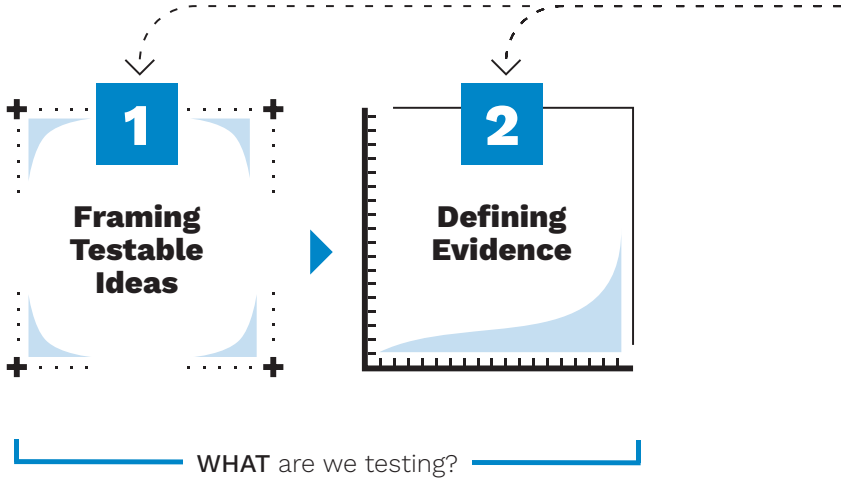
As we get started, a note to you, our reader: we wrote this book under the assumption that you have an idea of your own that you want to test. Great! This is the most effective way to work with this book—to learn while doing—whether it's with an actual idea you need to test or one you want to just use to learn with. We will use your ideas to practice the learning-by-doing process in Steps 1 through 5. In each step, we will give you an assignment to apply what we've talked about to your own idea—so you'll be both learning and doing as we go. We will ask you to note the milestones in your journey through the five steps using our **Progress Tracker** [TEMPLATE 1, PAGE 82].

By the time we conclude, you will have followed your idea through one full cycle of testing.

Let's get started exploring the world of experimentation!

TEMPLATE 1

Progress Tracker

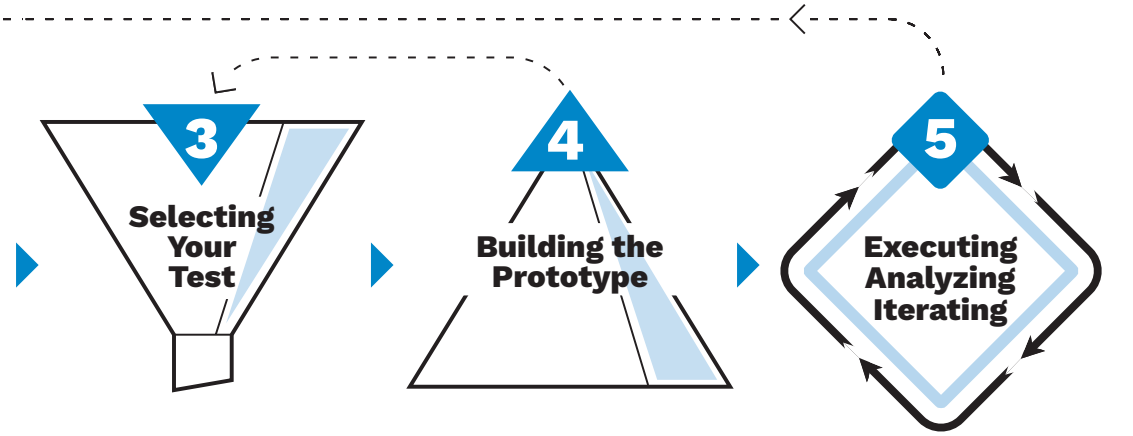


Step 1

- ☐ Frame a testable idea
- ☐ Prioritize concepts for testing
- ☐ Complete a Concept Snapshot

Step 2

- ☐ Surface critical assumptions
- ☐ Prioritize assumptions for testing
- ☐ Establish what constitutes evidence and identify targets



Step 3

- ☐ Choose between Say or Do tests
- ☐ Select a test type
- ☐ Design your test

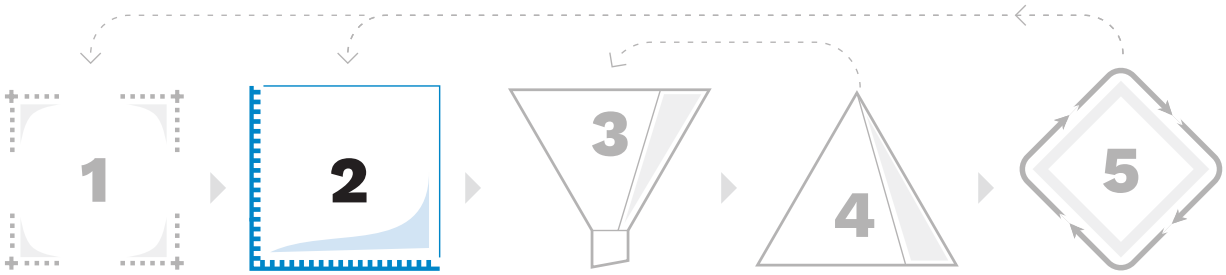
Step 4

- ☐ Assess the level of fidelity needed
- ☐ Select prototype format
- ☐ Construct your prototype

Step 5

- ☐ Audit and pre-test your design
- ☐ Run your test
- ☐ Share your findings
- ☐ Iterate your concept

Onto Step 1!



STEP 2

Defining Evidence

IN STEP 2 YOU WILL LEARN TO:

- Surface critical assumptions
- Prioritize assumptions for testing
- Establish what constitutes evidence and identify targets

WITH YOUR PRIORITIZED concept in hand, you are ready to define the kind of **evidence** you will look for to assess whether your idea is worth further investment. While "fail fast" has become a mantra for innovation efforts, this doesn't mean just tossing ideas into the market to see what happens. Doing quality experimentation requires careful design: If learning is your goal, you want to be sure that your failures are *intelligent* ones, ones that teach you something new and help you make tough choices among competing ideas. That means putting careful forethought into what you are testing for, what success looks like, and what data you need to collect to assess it. Here's where we start thinking about what specific data to collect and how to assess whether it supports or rejects our **hypothesis**. For example, Nike *thought* that people would return their shoes if it were easy for them to do so but wanted to gather evidence to support that **assumption**.

Definitions

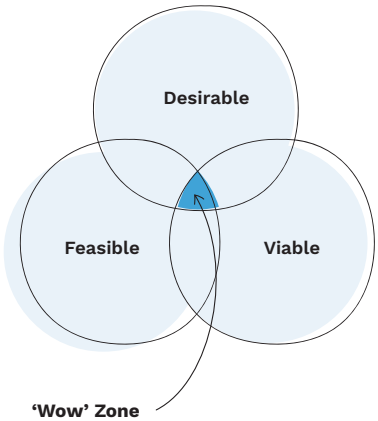
Assumptions	Underlying beliefs about why you think your idea is a good one
Evidence	Countable phenomena or data points that validate or refute a stated assumption—quantitative or qualitative
Hypothesis	The concept you want to test
Threshold target	The minimum value a metric needs to achieve to support moving the concept forward to the next stage of testing
Aspirational target	Future-focused target that describes the value you would like to see the metric achieve when your concept is successfully implemented in the real world
Source	Where you will find your evidence—may be archival (already existing) or new data gathered from the field

To define evidence, we:

1. **Surface the assumptions** behind any given concept;
2. **Prioritize these assumptions** to identify the most critical ones for testing; and
3. **Define the evidence** associated with them, its sources and **assign target metrics**, both threshold and aspirational. Sources may include both existing archival data and field data (to be gathered in the real world) and can be either qualitative or quantitative.

Sound complicated? It needn't be. We have found that a simple **fill-in-the-blank exercise**, can help to identify the different elements in this step.

Let's look at the sequence of activities in more detail. We begin by getting clear about what we are testing for. As we talked about earlier, we are looking for fresh ideas. They need not be new to the world—they just need to add value in three ways: **desirability** (be valuable to our target audience), **feasibility** (be able to be executed), and **viability** (be commercially sustainable over time). We call the intersection of these three conditions the Wow Zone:



Surfacing Assumptions Fill-in-the-Blank Exercise

One critical assumption I have about the ..... concept name ..... is that ..... assumption ..... .

prioritize assumptions

For this assumption to be true, I'd want to collect ..... quantitative / qualitative evidence ..... from ..... source ..... to confirm ..... hypothesis ..... .

define the evidence

At the end of this test, we need to see that ..... threshold target ..... to know we are on the right track towards our future aspirational target of ..... aspirational target ..... .

assign targets

Here's an example for SWR's *Concierge* concept:

One critical assumption I have about the ***Concierge* concept** is that **the concept creates a warm welcome**.

For this assumption to be true, I'd want to collect **comments on the service experience** from **customers who are intercepted for conversations** to confirm **adding *Concierge* at stations will build high-quality personal relationships with customers**.

At the end of this test, we need to see that **at least 50% of customers intercepted express a positive experience with the *Concierge* concept** to know we are on the right track towards our future aspirational target of **80%**.

Surfacing critical assumptions underlying your concept

To assess whether any given concept *wows*, we start by surfacing the assumptions we are making about *why* it meets each of the three conditions of desirability, feasibility, and viability. In other words, we articulate clearly why we believe that a concept belongs in the *Wow* Zone. These assumptions form the foundation of our testing strategy—they will point us towards the right metrics to gather that will signal whether or not these assumptions are valid.

You might be thinking, wait, I'm supposed to test my *assumptions*? I thought I was trying to test the concept itself? But in a hypothesis-driven problem-solving approach, you test the assumptions behind a concept rather than the concept itself. Why? Because our new concepts do not yet exist in the real world (only in our imagination), and we don't want to commit the time and resources to build them until we learn more about their desirability, feasibility, and viability. So instead, we test the assumptions underlying their attractiveness. At Whiteriver, they couldn't test the *Electronic Kiosk* itself without the time and expense of building it—but they could much more easily test the assumption that ER patients were comfortable using computers. **This testing of assumptions rather than ideas is how we manage our risk.** To test assumptions, we don't need to build a fully functional prototype—we just need a version functional *enough* to test the most critical assumptions underlying it. Without attention to assumptions, we risk investing in flawed ideas, jumping to conclusions, and falling victim to classic cognitive biases like the confirmation bias. Early on, testing the key assumptions underlying a concept, rather than the concept itself, is the fastest, cheapest way to learn.

Surfacing assumptions can turn out to be surprisingly difficult. Counterintuitively, it can be especially difficult for *experts* in different fields. In health care, for example, where the beliefs of highly trained clinical professionals have long driven decision-making, assumptions are often made about the needs and desires of patients that turn out not to be true.

Assumptions behind PMI's Spot Concept

Remember Spot, one of the four concepts that PMI wanted to test? Spot aimed to provide PMI members with real-world experiential learning by matching them with actual, low-risk opportunities to hone their skills. To test Spot, the Peer Insight team identified key assumptions under each of the three conditions:

For desirability, they wondered whether members would want to be involved in such projects. Would they see this as a good way to hone their skills?

For feasibility, they wondered if it would be possible to find the right kinds of projects and manage such a complex system of worldwide projects well?

And for viability, they worried that, even if it were feasible to find and run good projects, would the cost to do this be too high to make it viable long-term?

False Assumptions in Health Care

In Dallas, Children's Health Systems of Texas (CHST), one of the largest pediatric medical centers in the United States, faced clear challenges: the children they served faced some of the most troubling health indicators in the United States, with nearly 30 percent living in poverty. CHST leadership realized that their assumptions did not reflect the reality of their patients' lives. They initially assumed that patients and their families focused on preventative care with a proactive mindset, had strong support networks, and trusted caregivers' advice. The reality was that overwhelmed families experiencing poverty often struggled from crisis to crisis with a more reactive symptom-driven focus, lacked traditional support networks and were sometimes suspicious of caregivers' information. They needed a new approach based on patients' actual experiences, rather than their own beliefs as medical professionals.

In Melbourne, innovators at Monash University Medical Center wanted to train lay telecare guides to act as "professional neighbors" to keep in frequent telephone contact with elderly patients at high risk of hospital admission. They believed that carefully selected laypeople, trained in health literacy and empathy skills, and backed by decision support and professional coaches, could reduce hospitalization rates. Many of their clinical colleagues were skeptical, opposed to anyone other than a health professional performing such services, concerned about reducing the quality of care. Rather than debating this point, the innovators engaged their colleagues in the design of an experiment to test the value that lay telecare guides could deliver. Three hundred patients later, the results were in: overwhelmingly positive patient feedback and a demonstrated reduction in hospitalization rates and emergency room visits.

In Arizona, at Whiteriver Hospital, the Improvement Team realized that their assumption that ER patients were comfortable using computers was in error. Many tribal elders, the largest group of emergency room visitors, were not comfortable using the type of new technology that their planned Electronic Kiosk employed. Such a system, no matter how efficient it had proven to be in urban Baltimore, would likely create more, not fewer, delays at Whiteriver. As a result, the Team pivoted to a paper form aimed at identifying the severity of patients' medical issues as soon as they entered the emergency department. The simple form literally asked patients whether they needed emergency or nonemergency care, such as a visit with a nurse or a prescription refill, and anyone in the emergency room could help patients check boxes on a one-page form.

As you surface the assumptions behind your concept, you may find yourself tempted to just turn one of the conditions (like desirability) into a statement (Assumption: Users find our solution desirable). But at this abstract level, such an assumption is not testable. We need to push deeper to get more specific about why we believe that this particular condition is met (Assumption: Customers value the time savings that our product provides).

Surfacing assumptions works best as a team sport. Plus, bringing key stakeholders into agreement about what the important assumptions are and what it will take to confirm them is critical to achieving alignment. Structured conversations that put the right people in the room (real or virtual) to frame and plan the testing journey and collaboratively identify what success looks like will accelerate your progress by broadening your perspectives and building excitement, momentum, and alignment around what makes a concept a "wow." In all four examples we use in this field book, teams worked together to surface assumptions and determine what data they would need to test them.

In the earliest stages of experimentation, assumptions about desirability will be more important than assumptions about feasibility or viability. After all, what good is a feasible or viable solution that isn't desirable to the people who it was designed for?

We use the **Surfacing Assumptions** tool (TEMPLATE 5), organized by the three conditions we want to test for, to capture the relevant assumptions related to each of our prioritized concepts.

The team at SouthWestern Railway (SWR) working to test the *Concierge* concept, aimed at building high-quality personal relationships with customers, identified a set of assumptions related to **desirability** (creating a warm welcome that customers valued), **feasibility** (the service can run with existing staff levels), and **viability** (potential to self-fund).

TEMPLATE 5

SWR's *Concierge* Concept Assumptions

Desirability

- Creates a warm welcome that customers value
- Provides more efficient service for customers
- Increases SWR staff enjoyment of job

Feasibility

- Can flex to the needs of different stations
- Skills can be developed without costly training
- The service can be run with existing staff levels
- Existing assets can be employed

Viability

- Technology can be integrated to achieve self-service
- Potential increased retail space can be freed up
- Self-funding is possible

When you lay out the assumptions underlying your concepts, you will likely find that you have more assumptions than you can—or want—to test. Beware of letting the scope and cost of experiments balloon by trying to test too many assumptions. In Step 1, we prioritized our concepts to determine which ones to move into testing. Now, in Step 2, we will dig deeper and prioritize the critical assumptions underlying each concept. That will tell us where to start the testing process. Fortunately, not all assumptions are equally important—you start with the most critical ones, the ones that "make-or-break" your concept.

Prioritizing assumptions for testing

Almost all new concepts rest on a limited number of particularly significant assumptions. If these are true, they make the concept worth moving forward. If they are false, the rest of the assumptions don't matter enough to bother testing. To prioritize, focus your attention on two aspects of each of your assumptions:

1) How critical is that assumption to the success of the concept?
2) How much do you already know (from existing sources) about the likelihood that this assumption is true or false?

You can use the **Prioritizing Assumptions** tool (TEMPLATE 6) to capture this.

After putting together their list of assumptions for the *Concierge* concept, the SWR team prioritized their critical assumptions before proceeding.

Prioritizing Critical Assumptions behind the *Concierge* Concept

The DK&A team sorted the assumptions behind SWR's *Concierge* Concept, using the criteria of how important each was to the concept's success and what they already knew. Some, such as the possibility of self-funding and the use of existing assets, were considered to be low priorities for testing: they were not seen as critical, and quite a lot was known about them. Others, such as the training required to develop the new staff skills needed and the ability to integrate new technology, were seen as more critical, but SWR had confidence in their existing knowledge of them.

Three assumptions emerged as the most critical and unknown, all related to desirability.

The *Concierge* concept:

1. Created a warm welcome that customers valued,
2. Provided more efficient service, and
3. Increased SWR staff enjoyment of their jobs.

These three assumptions formed the focus of their initial testing efforts.

Translating assumptions into evidence

Once you have identified a small number of critical assumptions to prioritize for testing, we move from talking about what we *assume* to be true to talking about collecting *evidence* that it is true. This transition changes the conversation in two significant ways. First, it becomes personal. Assumptions reflect properties of the new concept and should be visible to all—but whether something is "proven" true lies in the eyes of the beholder. Pay attention to *who* needs to conclude that something is true and how they see the world. We tailor the identification of metrics to the key stakeholders we must convince. Invest some time in thinking about who your key stakeholders are and what their relationship is to the concept you want to test. Who needs convincing? How skeptical are they? What is at stake for them? How fast do they expect you to generate evidence?

You can use these questions as an exercise in co-creation—invite a diverse set of your key stakeholders into a conversation aimed at answering them. At SWR, for example, one of the key stakeholder groups, senior SWR leadership, were anxious to see credible test results in actual station environments as soon as possible.

Warning

You may be tempted to skip the surfacing of assumptions and go straight to identifying evidence, based on the data you have available. This is a serious mistake, as it will encourage you to work backward from the data you've got. That is not hypothesis driven! Instead, you want to start with the assumptions you need to test and only then specify the right kind of evidence to assess them (which may or may not already exist).

The second difference, as we move toward collecting evidence, is the level of specificity. Though we have worked hard to make our assumptions less ambiguous, evidence must be observable and countable, a more demanding standard. It must also fit the context of the test you are about to do. Some issues to consider as you define your evidence:

- How much time do you have available for the testing process?
- What kinds of resources can you call on?
- How big is your budget? What can you afford?
- What is possible from a technical viewpoint?

The DK&A team asked themselves how they would know if the three critical assumptions behind the *Concierge* concept were true. How would these show up in observable and countable ways? To test the creates a warm welcome assumption, they considered evidence based on how many customers sought out the *Concierge* services and whether they expressed appreciation for the services. For the increased efficiency assumption, they looked for reduced wait times at windows and increased use of self-service. For improved employee satisfaction, they decided to solicit staff members' feelings about their new role.

The form that evidence takes often needs to be iterated as you push further into the experimentation process and learn more. In general, as projects progress, the desired

evidence increases in terms of specificity, becomes more quantitative, and comes from multiple sources to manage the risk of the increasing investment being made in the new concept. In the early stages of testing, it is not always obvious what to measure. Data sources in the real world are messy, and more than one metric is often needed to show whether assumptions are true or not. Sometimes the best we can do is to verify whether we are *directionally* correct. For these reasons, triangulation from multiple data sources is always valuable.

It makes sense to begin your search for the right evidence by reviewing those measures already in use that may have value for your investigation and ask yourself the following questions: What evidence has meaning in this sector, with the audience you must convince, or for this type of activity? What is already being measured? Can you adapt ongoing monitoring to support your experiment? Examining existing data is an important step in preparing to go into the field - just make sure to remain hypothesis driven!

The relative value of *qualitative* versus *quantitative* data is also important to consider. Quantitative data may be already available, but do they measure what you need? Qualitative data, the kind that is good for sense-making and that assures directional correctness, may be essential, but needs to be gathered from the field. In addition, some audiences may prefer quantitative over qualitative data, while others love stories.

TEMPLATE 7

Assumptions to Evidence for SWR's *Concierge* Concept

Assumption	Evidence	Source
Creates a warm welcome that customers value	Frequency of customers seeking interactions with *Concierge* staff	Observation in ticket hall
	Positive comments from customers expressing appreciation for the service	Customer intercept interviews
Provides more efficient service for customers and enables improved interaction with SWR staff	Volume/speed/accuracy of *Concierge* staff answers	Observation in ticket hall
	Length of queues at ticket windows	Observation in ticket hall compared with historical record
	Increased volume of tickets purchased at self-service window	Sales data from self-service machines compared with historical record
Increases SWR staff enjoyment of job	Staff appreciation for new role and interest in remaining in new role versus return to ticket window	Staff interviews

As you specify the evidence for your concepts, you will likely want a combination of qualitative and quantitative data, drawn from both archival and field sources, at each stage. Once these are specified, you need to identify the **source** of the data. Ask yourself *where* you will find the data you need.

The DK&A team, for instance, identified evidence in support of the *Concierge* concept that was both qualitative (e.g., customer and staff satisfaction) and quantitative (e.g., queues at ticket windows, sales data from automated ticket machines).

As a final activity in Step 2, it is valuable to identify and differentiate between **aspirational targets** (the ones that you hope your idea will eventually produce) and **threshold targets** (which tell you whether to go to the next step) and offer both for each metric. It is unrealistic to believe that you will reach aspirational targets over the course of hours, days, or weeks. Even so, setting a target up front remains a useful activity that provides information to assess the magnitude of change and how long it might take to reach the aspirational targets. You can then consider whether that aligns with your project timeline to reach the desired impact or not. If yes, carry on. If not, you may need to go back to Step 1 and select a different concept for testing.

Tips for Surfacing Assumptions

- **What questions are still outstanding?** Ask yourself, "What don't I know for certain?"
- **Put on an investor's hat:** What questions would need to be answered to receive another round of funding? What do you need to prove?
- **Convert your questions into affirmative statements** (e.g., Customers want to recycle old shoes).
- **Think about the categories:** desirability, feasibility, viability.
- **Be sure your assumption is specific enough.**
- **Revisit some of the design tools** traditionally used to inspire idea generation, like journey maps, jobs-to-be-done, personas, or value chain maps, to help you surface assumptions.

Tips for Identifying Targets

For each of the quantitative evidence, ask yourself: what is the smallest amount of change you'd want to see in order to feel solid about moving this concept forward for further testing? This will give you your threshold target. Then ask yourself: what is the desired/hoped for amount of change when the concept is successfully implemented? This helps pinpoint your aspirational target.

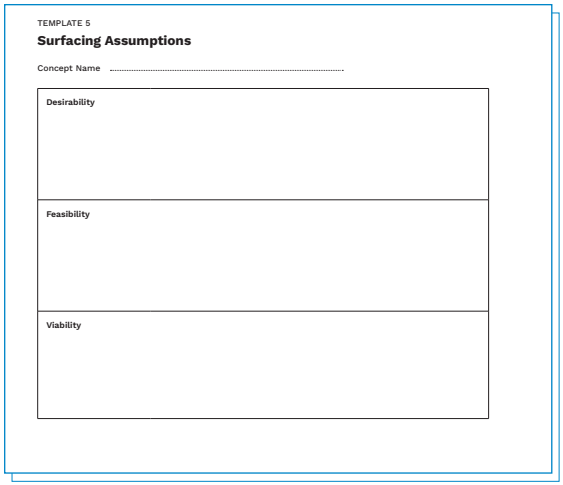
For each qualitative evidence, ask yourself: what are the near-term expected responses that would signal that you are headed in the right direction? These will give you a threshold target. Then ask: what are the desired/hoped-for responses you would want to see in the future? This signifies your aspirational target.

Assignment

It's assignment time again! (Remember that you can find the full-scale templates beginning on *page 82*.)

1

Use the **Surfacing Assumptions** tool [TEMPLATE 5, *page 87*] to lay out the key assumptions you believe underlie the concept you have selected for testing, according to the conditions of desirability, feasibility, and viability.


TEMPLATE 5

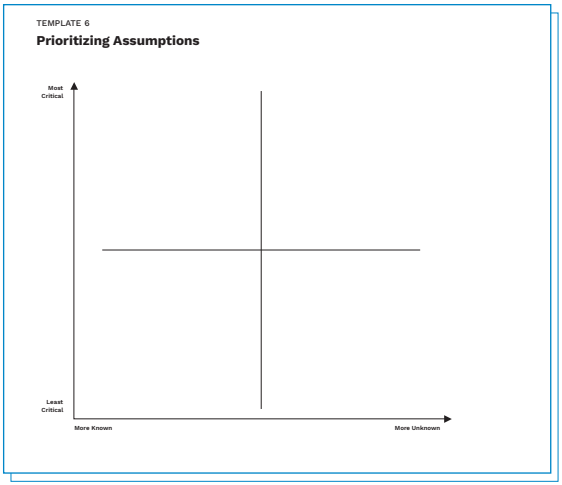
Surfacing Assumptions

Concept Name .....

Desirability
Feasibility
Viability

2

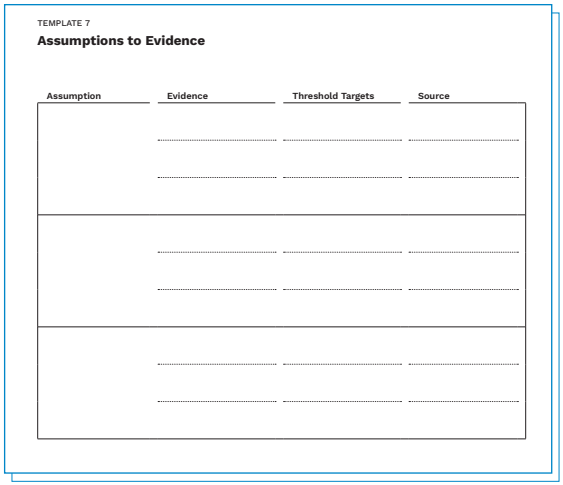
Next, identify which assumptions are most critical, and decide which ones it is most important to test first using the **Prioritizing Assumptions** tool [TEMPLATE 6, *page 88*].



ASSIGNMENT

3

Finally, using the **Assumptions to Evidence** tool [TEMPLATE 7, *page 89*], capture the evidence, sources, and targets associated with each assumption.



TEMPLATE 7

Assumptions to Evidence

Assumption	Evidence	Threshold Targets	Source

Log your progress in your **Progress Tracker** [TEMPLATE 1, *page 82*]:

- ☑ Surface critical assumptions
- ☑ Prioritize assumptions for testing
- ☑ Establish what constitutes evidence and identify targets

Now that you have a clear set of assumptions, metrics, and targets for the concept you want to test, you are ready to move from the *what* of testing to the *how*. Then move onto the next step in the process—selecting a type of test.

Onto Step 3!